



Universidade de Brasília

FACULDADE UnB PLANALTINA

LICENCIATURA EM CIÊNCIAS NATURAIS

**ANÁLISE DA QUANTIDADE DE *MISSING DATA*
COMO CRITÉRIO PARA INCLUSÃO OU EXCLUSÃO
DE TÁXONS FÓSSEIS EM ANÁLISES
FILOGENÉTICAS**

ANDREZZA ROMENIA LIMA DE ABREU

ORIENTADOR: Prof. Dr. RODRIGO MILONI SANTUCCI

Planaltina – DF

Dezembro 2013



Universidade de Brasília

FACULDADE UnB PLANALTINA

LICENCIATURA EM CIÊNCIAS NATURAIS

**ANÁLISE DA QUANTIDADE DE *MISSING DATA*
COMO CRITÉRIO PARA INCLUSÃO OU EXCLUSÃO
DE TÁXONS FÓSSEIS EM ANÁLISES
FILOGENÉTICAS**

ANDREZZA ROMENIA LIMA DE ABREU

ORIENTADOR: Prof. Dr. RODRIGO MILONI SANTUCCI

Trabalho de Conclusão de Curso apresentado à Banca Examinadora, como exigência parcial para a obtenção de título de Licenciado do Curso de Licenciatura em Ciências Naturais, da Faculdade UnB Planaltina, sob a orientação do Prof. Dr. Rodrigo Miloni Santucci.

Planaltina - DF

Dezembro 2013

AGRADECIMENTOS

Agradeço a Deus primeiramente, por todas as bênçãos recebidas, por ser minha fortaleza e meu refúgio, e ter me dado forças para prosseguir sempre, e alcançar mais essa vitória em minha vida.

A minha família, especialmente a minha mãe, por ser meu maior exemplo de determinação e perseverança, e por todo amor e apoio.

A todos os queridos amigos que conheci na FUP, e que com certeza levarei para sempre no meu coração, muito obrigado por todos os momentos de aprendizagem, alegrias e tristezas compartilhados, as viagens, as noites de estudos em claro, sem vocês o curso não teria sido tão maravilhoso. Especialmente ao meu trevo querido, que jamais esquecerei.

Aos brilhantes professores da FUP, que com seus excelentes ensinamentos foram fundamentais em minha formação. As professoras Alice Melo Ribeiro e Anete Maria de Oliveira, minhas queridas orientadoras de Iniciação Científica e Prodocência, muito obrigada pela oportunidade de trabalhar com vocês e por todos os ensinamentos essenciais para meu amadurecimento acadêmico.

As minhas amigas de estágio da ETSP, muito obrigada pela compreensão, força, apoio e palavras de incentivo, por não me deixarem desanimar.

Ao Roger pela criação do programa GMMD, essencial para este trabalho.

Agradeço, ESPECIALMENTE, ao professor Rodrigo Miloni Santucci, meu orientador, por me acompanhar nessa etapa, por todo apoio, paciência e ensinamentos fundamentais para desenvolvimento deste trabalho.

ANÁLISE DA QUANTIDADE DE *MISSING DATA* COMO CRITÉRIO PARA INCLUSÃO OU EXCLUSÃO DE TÁXONS FÓSSEIS EM ANÁLISES FILOGENÉTICAS

Andrezza Romenia Lima de Abreu¹

RESUMO

A Sistemática Filogenética revela um padrão de agrupamento dos organismos baseado em sinapomorfias. A matriz de dados apresenta as informações de determinado táxon em relação a um caráter ou estado de caráter. A utilização de táxons incompletos em análises filogenéticas pode levar a um problema na falta de resolução da análise. Isso porque táxons ou caracteres incompletos em uma matriz de dados podem resultar em inúmeros cladogramas igualmente mais parcimoniosos. Este problema atinge principalmente táxons fósseis, por apresentarem em sua matriz de dados grande quantidade de *missing data*, que são simplesmente informações desconhecidas. A quantidade de *missing data* pode resultar na exclusão de um táxon fóssil da análise. Mas de fato existe um limite mínimo de *missing data* que um táxon pode apresentar para sua inclusão ou exclusão da análise? Este trabalho busca estudar táxons incompletos com grande quantidade de *missing data* e analisar se a falta de resolução da árvore evolutiva resultante é causada pela quantidade global de informações perdidas ou por *missing data* que representam especificamente sinapomorfias. Foram analisadas 250 matrizes de dados geradas aleatoriamente contendo diferentes porcentagens de *missing data* e 10 casos especiais para o comportamento de um determinado táxon. Os resultados mostram que os *missing data* não apenas causam politomias nas árvores, quando se referem à sinapomorfias excluídas, como também modificam a topologia das árvores, o que dificulta o entendimento e a interpretação da evolução, bem como a compreensão da diversidade biológica.

Palavras-chave: *missing data*; análise filogenética; táxons fósseis; sinapomorfias.

1. INTRODUÇÃO

A Sistemática Filogenética possui dois objetivos principais, um deles é fornecer subsídios para uma compreensão geral da diversidade biológica, da evolução dos táxons e da modificação de caracteres; o outro está relacionado em desenvolver a habilidade de propor hipóteses sobre a evolução de caracteres ou sobre as relações de parentesco entre os membros de um grupo (AMORIM, 2002). A Sistemática Filogenética agrupa os organismos com base em características derivadas, compartilhadas e exclusivas, ou seja, com base em sinapomorfias. Isso porque, são a única evidência de uma origem evolutiva comum entre os organismos, e somente os cladogramas podem identificar essas homologias. Hennig (1966, p. 20 apud AMORIM, 2002, p. 15) salienta que:

Provavelmente, tem grande significado histórico o fato de o próprio Darwin ter dito que a possibilidade de ordenar os organismos em um sistema hierárquico só é explicável assumindo-se uma relação filogenética entre eles: “o simples fato de que as espécies, tanto extintas quanto viventes, agrupam-se em gêneros, famílias, ordens etc. – uma divisão análoga à variabilidade subjacente – não seria explicável de outra maneira, e pode não nos parecer extraordinário apenas por ser um lugar-comum.

Segundo Amorim (2002), o método de reconstrução filogenético é um sistema para identificar sinapomorfias, e assim delimitar grupos monofiléticos (um conjunto de táxons

¹ Licenciatura em Ciências Naturais - Faculdade UnB de Planaltina

incluindo um ancestral comum exclusivo e todas as suas espécies descendentes). Essa reconstrução é feita através de matrizes de caracteres, as quais são verdadeiros depósitos de informações sobre nosso conhecimento a respeito da diversidade biológica. No cladograma as sinapomorfias podem se apresentar de forma ambígua, quando, por exemplo, alguns táxons dentro de um suposto grupo monofilético não possuem essa informação preservada, ou seja, ela pode ser otimizada como presente ou ausente. Por outro lado, quando todos os táxons dentro de um grupo monofilético apresentam essa característica preservada temos uma sinapomorfia não ambígua.

As matrizes exibem a condição de um caráter em um determinado táxon (qualquer classe cujos elementos são organismos reunidos com base em semelhanças, seja uma espécie, uma parte de uma espécie ou um agrupamento de espécies). “Não são a fonte primária de informação sistemática – os próprios organismos são as fontes primárias –, mas correspondem à base de dados sobre a qual os cladogramas são construídos” (AMORIM, 2002, p. 45).

O cladograma é um dendrograma (diagrama ramificado) que expressa relações filogenéticas apenas entre táxons terminais (Figura 1), evidenciadas por sinapomorfias, podendo ocorrer também homoplasias, que são semelhanças adquiridas independentemente em dois ou mais táxons (AMORIM, 2002).

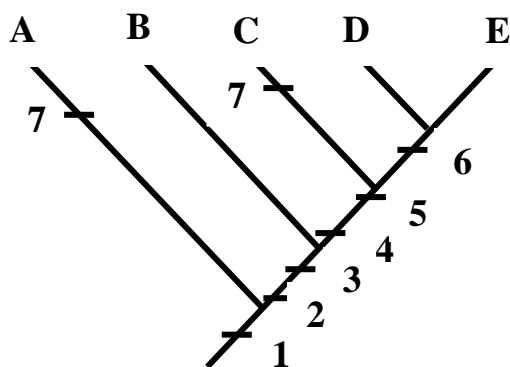


Figura 1: Exemplo de cladograma mostrando as relações de parentesco entre os táxons A até E e as respectivas sinapomorfias (características derivadas, compartilhadas e exclusivas) e homoplasias (caráter 7) que agrupam esses táxons.

No cladograma, as conexões entre as espécies indicam apenas que há uma história evolutiva, ou descendência, comum, mas não aponta uma relação ancestral-descendente. Já uma Árvore Evolutiva segundo Amorim (2002) é um dendrograma que expressa relações filogenéticas tanto entre táxons terminais, quanto entre espécies ancestrais e espécies descendentes, mas não se utiliza apenas de sinapomorfias para estabelecer as relações evolutivas entre os táxons.

Hipóteses de homologia são a base da Sistemática Filogenética, mas homologias não podem ser avaliadas para partes desconhecidas de organismos. Os *missing data* são reconhecidos como um problema em Paleontologia porque as relações evolutivas dos táxons para os quais poucas hipóteses de homologia podem ser estabelecidas são, às vezes, difíceis de resolver. Mas este problema não está limitado a táxons fósseis, nem necessariamente são mais problemáticos com eles. Dessa forma, podemos encontrar nas matrizes que envolvem táxons fósseis uma grande quantidade de *missing data*, que representam informações desconhecidas, ou a ausência de informação. Quando um táxon incompleto é incorporado em uma matriz de dados com outros táxons, pode aparecer um grande número de cladogramas

equivalentes, onde o consenso estrito destes pode apresentar pouca resolução. Geralmente este problema é atribuído aos *missing data*, e, portanto, devem ser evitados, e uma possível solução é a exclusão de táxons incompletos na análise (KEARNEY e CLARK, 2003).

Vários pesquisadores têm utilizado a abordagem de exclusão de táxons fósseis de acordo com o percentual de *missing data* que representam na matriz de dados (Rowe, 1988; Fraser e Benton, 1989; Benton, 1990; Gao e Norell, 1998; Grande e Bemis, 1998). Esta abordagem pressupõe que quando táxons incompletos estão incluídos na análise isso pode resultar no aumento de ambiguidade, ou seja, que a proporção de *missing data* está correlacionada com o grau de ambiguidade da análise.

Trabalhos como o de Fraser e Benton (1989) apontam, entretanto, que a resolução da análise não aumenta com a progressiva exclusão de táxons ou caracteres incompletos da matriz de dados. Eles excluíram sucessivamente táxons com base no percentual de *missing data* em sua análise de esfenodontes e ilustraram os resultados para cada uma das referidas análises na figura abaixo (Fig. 2):

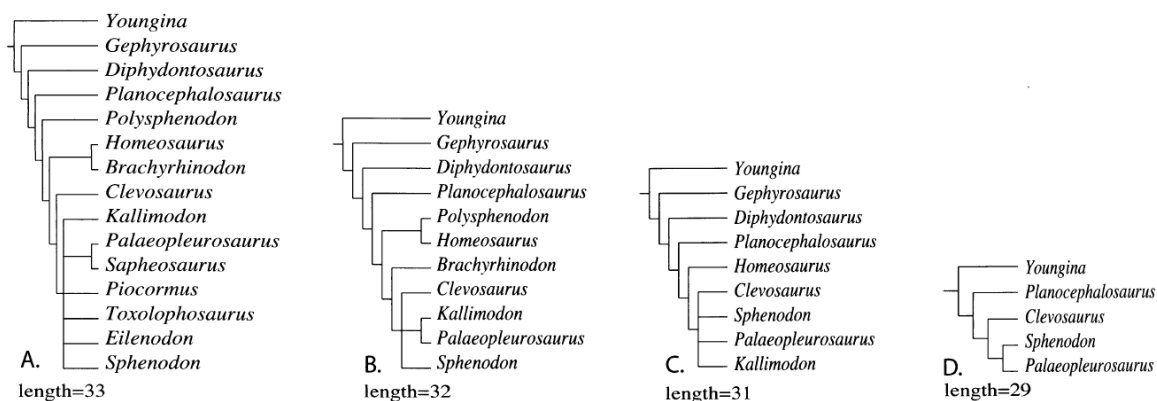


Figura 2: A, consenso estrito de 82 cladogramas mais parcimoniosos com base na análise incluindo todos os táxons. B, consenso estrito de 2 cladogramas mais parcimoniosos excluindo 4 táxons com mais de 50% de *missing data*. C, consenso estrito de 4 cladogramas mais parcimoniosos (todos os táxons com mais de 80% de *missing data* foram excluídos). D, 1 cladograma mais parcimonioso incluindo somente táxons completos (0% de *missing data*) da matriz. Fonte: Fraser e Benton, 1989.

Sua análise demonstra que esta estratégia é imprevisível, os números das árvores não diminuíram de forma linear. Conforme táxons incompletos são excluídos, mais árvores foram produzidas na análise 3 (Fig. 2C) do que na análise 2 (Fig. 2B), com um aumento na ambiguidade, apesar de uma diminuição de *missing data* neste caso particular. Além disso, a análise final resultou em um único cladograma com apenas 5 táxons completos, dos 15 originais, de forma que pouco poderia ser concluído quanto à questão original da análise. Dessa forma, usar a quantidade de caracteres completos como critério para inclusão ou exclusão de um táxon não é aconselhável (KEARNEY e CLARK, 2003).

Adicionalmente, toda evidência relacionada a uma hipótese deve ser considerada, e a exclusão de evidências que poderiam tanto suportar ou refutar uma hipótese é difícil de ser justificada. Entretanto, a inclusão de táxons incompletos pode levar a um problema operacional de múltiplos cladogramas igualmente mais parcimoniosos com uma concomitante ambiguidade nos resultados. “Os *missing data* simplesmente representam o desconhecido e não devem ser vistos como um impedimento para a consideração de toda evidência disponível em uma análise filogenética, nem devem servir de justificativa para exclusão de táxons ou

caracteres.” (KEARNEY e CLARK, 2003, p. 263). Assim, parece haver algum efeito adicional que afeta a resolução dos estudos cladísticos, além da pura e simples quantidade de *missing data* presente na análise original.

Esse trabalho tem como objetivo estudar os táxons incompletos (com grande quantidade de *missing data*) em análises filogenéticas, e verificar se a falta de resolução da árvore evolutiva é causada pela quantidade global de informações perdidas ou se é causada especificamente por *missing data* que representam sinapomorfias. Dessa forma, iremos analisar as condições para inclusão ou exclusão de um táxon fóssil em análises filogenéticas, verificando a relação entre a porcentagem de *missing data* e o número de árvores mais parcimoniosas e testar se existe um limite mínimo de *missing data* para exclusão de um táxon ou caráter.

2. METODOLOGIA

O trabalho de pesquisa foi dividido por etapas. Etapa 1 – Inicialmente foi realizado um estudo bibliográfico a respeito do tema, a fim de conhecer as hipóteses existentes. Em seguida, utilizando o programa GMMD – Gerador de matrizes e *missing data* (<http://www.sindicatodosbordoos.com.br/gmmd/index.php>) desenvolvido para este trabalho, o qual gerou randomicamente 25 matrizes contendo um conjunto de 10 táxons e 100 caracteres cada, e 25 matrizes contendo um conjunto de 20 táxons e 100 caracteres cada, sendo consideradas como matrizes originais. Essas matrizes iniciais são completas, todos os táxons e caracteres são codificados, estas foram consideradas como as matrizes originais.

A escolha da quantidade de táxons foi baseada no estudo de Harcourt-Brown et al. (2001), que compararam árvores paleontológicas com árvores neontológicas utilizando o índice de Colless's, o qual demonstra estatisticamente a tendência ao não balanceamento das árvores quando este índice é mais próximo do valor 1, e mais balanceada a árvore quando este é mais próximo de 0. Há diferentes tipos de balanceamento de árvores (Fig. 3).

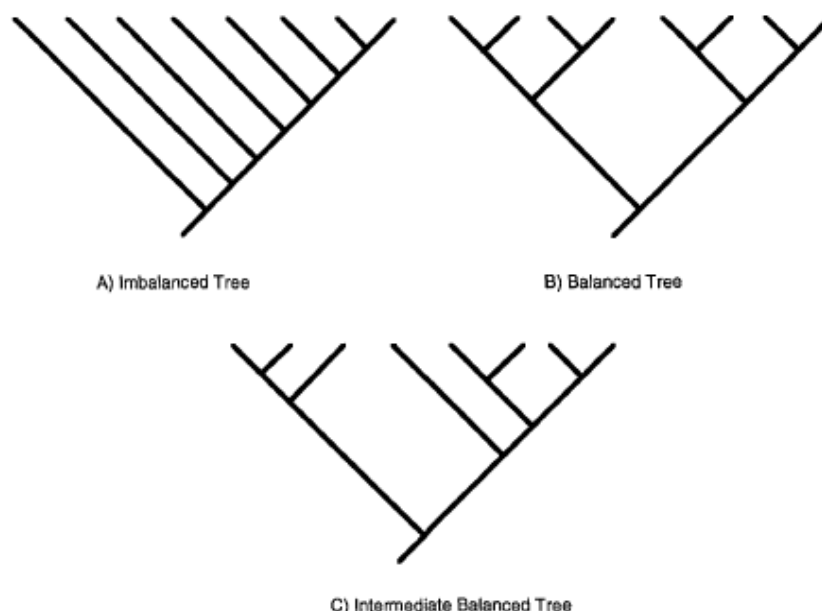


Figura 3: Diferentes tipos de balanceamento de árvores, A - exemplo de uma árvore não balanceada, B - exemplo de uma árvore balanceada e C - exemplo de uma árvore com balanceamento intermediário. Fonte: Harcourt-Brown et al., 2001.

Segundo Harcourt-Brown et al. (2001), o balanceamento intermediário é o mais comum, para que ocorra uma árvore balanceada é necessário que o número de táxons seja par. Porém, para que haja um maior balanceamento é necessário um número maior de táxons, isto é, quanto maior o número de táxons, maior será a aproximação do valor zero no índice de Colless's, logo maior as chances de se obter uma árvore mais balanceada (Fig. 4).

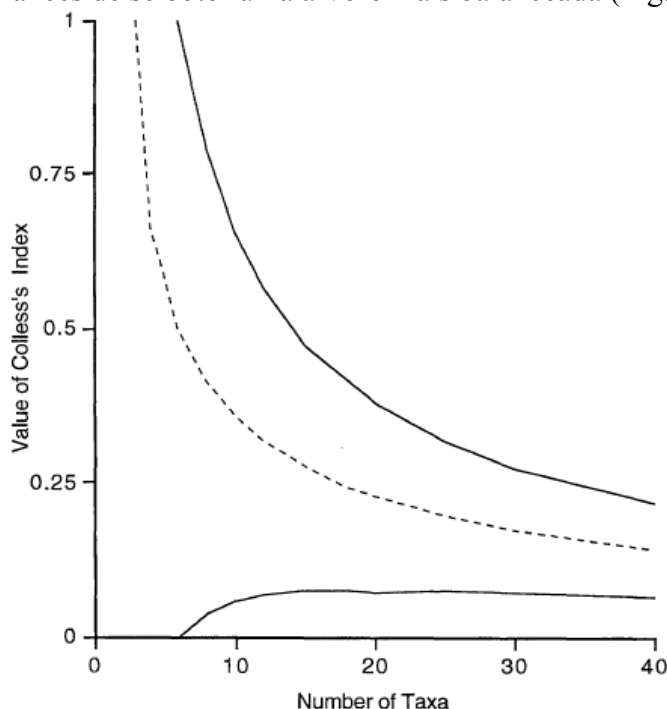


Figura 4: Representação da relação entre o número de táxons e o índice de Colless's. Os dados para este modelo foram obtidos a partir do modelo de Rogers 1994. O modelo é representado por uma linha tracejada para o valor esperado e por linhas contínuas para dois desvios padrões. A partir de 20 táxons há uma tendência de se obter árvores mais balanceadas. Fonte: Harcourt-Brown et al. (2001).

De acordo com o gráfico acima, pode-se inferir que quanto maior o número de táxons, maior a possibilidade de obtermos árvores balanceadas, pois esse valor se aproxima mais de zero. O contrario ocorre à medida que diminuimos a quantidade de táxons, o valor do índice se aproxima mais de um, logo obteremos árvores com um menor balanceamento ou não balanceadas. A quantidade de táxons (10 e 20) foi pensada de maneira que se avaliasse o efeito dos *missing data* em árvores balanceadas e não balanceadas.

A quantidade utilizada de caracteres (100) foi escolhida por apresentar uma proporção razoável para o número de táxons: 1 para 5, no caso de 20 táxons e 1 para 10, no caso de 10 táxons, além de facilitar a transformação do número de *missing data* em porcentagem.

Etapa 2 - A partir das matrizes geradas na etapa anterior foram construídas árvores filogenéticas por meio do programa computacional PAUP (*Phylogenetic Analysis Using Parsimony*) versão 4.0b8, Sinauer, Sunderland, Massachusetts (SWOFFORD, 2001). Onde inicialmente enraizamos todas as matriz ao táxon 1, em seguida colapsamos todas as árvores de tamanho zero e processamos as matrizes para 10000 replicatas cada, usando o método de adição randômico e a opção de *swapping* TBR para se evitar ótimos locais. Este processo foi

bem demorado, pois as matrizes originais selecionadas deveriam conter o mínimo possível de politomias iniciais, isso para todas as matrizes originais, e o consenso estrito foi utilizado quando a matriz gerada apresentava mais de 1 árvore mais parcimoniosa. Se o consenso estrito apresentasse um grande número de politomias, ou por exemplo, uma politomia basal, esta matriz não era utilizada e seguíamos em busca de novas matrizes que apresentassem em suas árvores uma topologia sem politomias.

Etapa 3 - Esta etapa consiste na exclusão aleatória de alguns estados de caráter para alguns táxons das matrizes originais, gerando certa quantidade de *missing data*. A partir de cada matriz original foram geradas 4 novas matrizes, contendo *missing data*. A primeira matriz foi considerada como matriz original contendo todas as informações, logo sendo uma matriz completa, a segunda matriz derivada da matriz original contém 90 % de informações completas, ou seja, essa matriz possui 10 % de *missing data*, a terceira matriz também derivada da matriz original possui 75 % de informações, ou seja, essa matriz possui 25 % de *missing data*, a quarta matriz derivada da matriz original possui 50 % de informações e por consequência 50% de *missing data*, e por fim a quinta matriz derivada da matriz original possui 25 % de informações e 75% de *missing data*. Temos então um total de 250 matrizes incluindo as matrizes originais e as matrizes com as diferentes porcentagens de *missing data*, sendo 125 matrizes contendo 10 táxons e 100 caracteres e 125 matrizes contendo 20 táxons e 100 caracteres.

Etapa 4 – Depois de geradas as 250 matrizes começamos uma análise para cada matriz, incluindo as matrizes com as diferentes porcentagens de *missing data*. Todas estas matrizes foram processadas no PAUP e utilizamos o comando *describetrees n/apolist*, na qual verificamos em cada matriz o número de nós (NN), número de árvores mais parcimoniosas (MPT), número de passos (PASSOS), número de sinapomorfias ambíguas (NSA) e número de sinapomorfias não ambíguas (NSNA). Todos os gráficos e planilhas desta análise estão disponíveis nos anexos deste trabalho.

Outra análise foi realizada para verificarmos o comportamento de um determinado táxon e/ou seu grupo-irmão, “que corresponde ao grupo monofilético mais próximo de um determinado táxon” (RIOS-NETO, 2010, p. 183). Nas matrizes contendo 25% de informação com relação a sua matriz original com 100% de informação, utilizamos o PAUP para processar a matriz e, nesse caso, usamos o consenso de maioria para detectar os grupos irmãos quando não foi possível usar o consenso estrito, devido às politomias. Em seguida com o comando *describetrees n/apolist*, verificamos a quantidade de sinapomorfias ambíguas e não ambíguas para este agrupamento, por fim verificamos nas matrizes de 25% quantos caracteres foram excluídos e a topologia das árvores em relação à original. Esse procedimento, devido ao grande tempo demandado, pois se trata de uma análise manual, foi repetido para 5 matrizes escolhidas aleatoriamente.

3. RESULTADOS E DISCUSSÕES

Abaixo um exemplo do gráfico da análise para a matriz 1 das matrizes com 10 táxons e 100 caracteres (Gráfico 1):

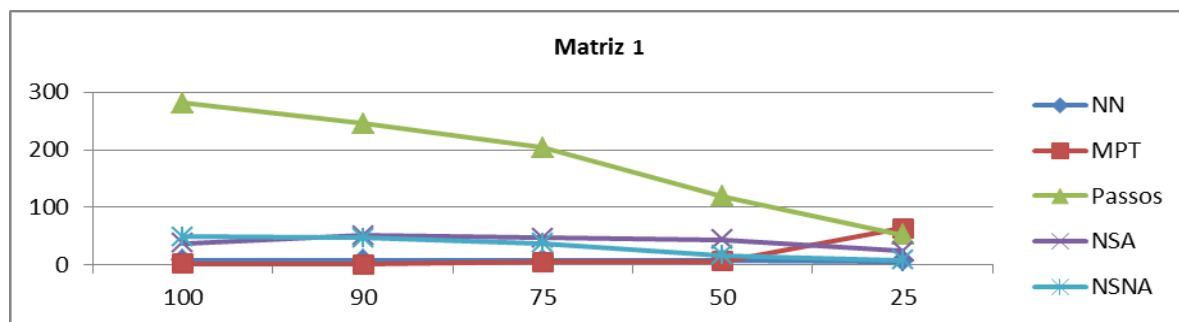


Gráfico 1: Exemplo de gráfico gerado a partir da matriz original e suas respectivas reduções (matriz de 10 táxons e 100 caracteres). No eixo x encontram-se as porcentagens de informação para cada versão da matriz, variando de completa (100%), até com 25% de informação presente, a qual contém 75% de *missing data*. O eixo y representa os valores de cada parâmetro analisado.

Neste gráfico podemos verificar que o NN se mantém praticamente constante em todas as árvores da matriz 1, o MPT se mantém aproximadamente constante para todas as árvores com exceção da árvore de 25%, onde podemos ver uma elevação no número de MPTs. Este aumento de MPTs ocorre nas matrizes 1, 4, 6, 9, 11, 13, 14, 16, 18, 19, 21, 24 e 25. As matrizes 23, 22, 13, 10, 8, 7, 5 apresentaram uma pequena elevação de MPT, o restante manteve-se constante. O número de passos diminui continuamente de acordo com a quantidade de *missing data*, ou seja, a árvore original (100%) possui mais passos que a árvore de 75%, que consequentemente possui mais passos que a de 50% e assim sucessivamente, e isso ocorre em todas as matrizes. O NSA é aproximadamente o mesmo para todas as matrizes, possui um leve aumento entre as árvores de 90 e 50%, depois cai um pouco nas de 25%. O NSNA para todas as matrizes se comporta aproximadamente igual, inicia com um número maior nas árvores de 100% e esse número vai diminuindo à medida que se aproxima da árvore de 25%.

Para as matrizes de 20 táxons e 100 caracteres temos um exemplo abaixo (Gráfico 2), no qual obtivemos os seguintes resultados:

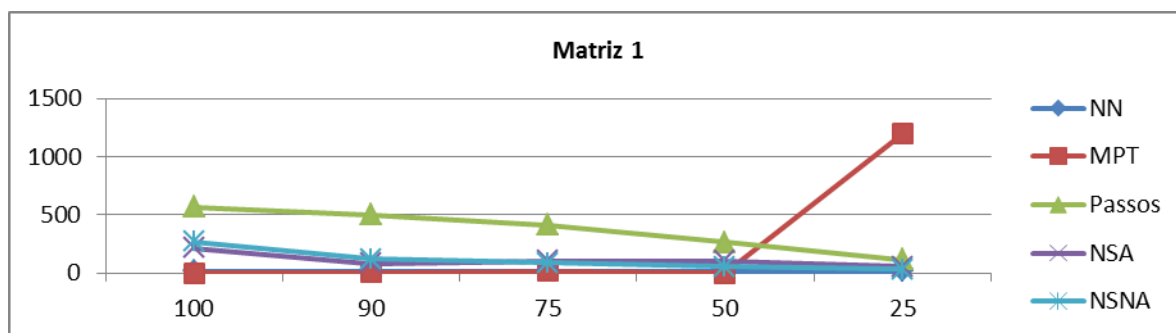


Gráfico 2: Exemplo de gráfico gerado a partir da matriz original e suas respectivas reduções (matriz de 20 táxons e 100 caracteres). No eixo x encontram-se as porcentagens de informação para cada versão da matriz, variando de completa (100%), até com 25% de informação presente, a qual contém 75% de *missing data*. O eixo y representa os valores de cada parâmetro analisado.

De acordo com o gráfico, podemos perceber que o NN, assim como nas matrizes de 10 táxons, se mantém aproximadamente constante para todas as árvores, o MPT na maioria das matrizes se mantém constante até a árvore de 50% e depois tem um aumento considerável

nas árvores de 25%, com exceção da matriz 2 e 5 que tem um pequeno aumento desde a árvore de 50%. As matrizes 7 e 12 também revelam um leve aumento na árvore de 50%, mas há uma queda nesse número na árvore de 25%. Nas matrizes 16 e 18 há um leve aumento das MPTs na árvore de 75% e depois se estabiliza na árvore de 50% e, por fim, tem um aumento novamente nas árvores de 25%. Nas matrizes 17 e 19 as MPTs se mantêm constantes e iguais para todas as árvores da matriz, não há aumento nem diminuição do número de MPTs. O número de passos diminui conforme o aumento de *missing data*, assim como ocorre nas matrizes de 10 táxons. O NSA se mantém aproximadamente constante, mas há uma leve oscilação nas matrizes, o NSNA também se mantém aproximadamente constante em todas as matrizes, assim como nas matrizes com 10 táxons, e vai diminuindo principalmente nas árvores de 25%.

Quanto a análise do comportamento de um determinado táxon e/ou seu grupo irmão obtivemos o seguinte resultado para matrizes de 10 táxons e 100 caracteres:

Matriz 2 de 10 táxons/100 caracteres 25% (Figura 5)

Grupo-irmão: 8 e 9

Quantidade de sinapomorfias no ramo 15>14 = 0 ambígua e 1 não ambígua.

Quantidade de caracteres excluídos no táxon 8 = 71 caracteres.

Quantidade de caracteres excluídos no táxon 9 = 81 caracteres.

```
Tree length = 60
Consistency index <CI> = 0.9333
Homoplasy index <HI> = 0.0667
CI excluding uninformative characters = 0.7778
HI excluding uninformative characters = 0.2222
Retention index <RI> = 0.7333
Rescaled consistency index <RC> = 0.6844
```

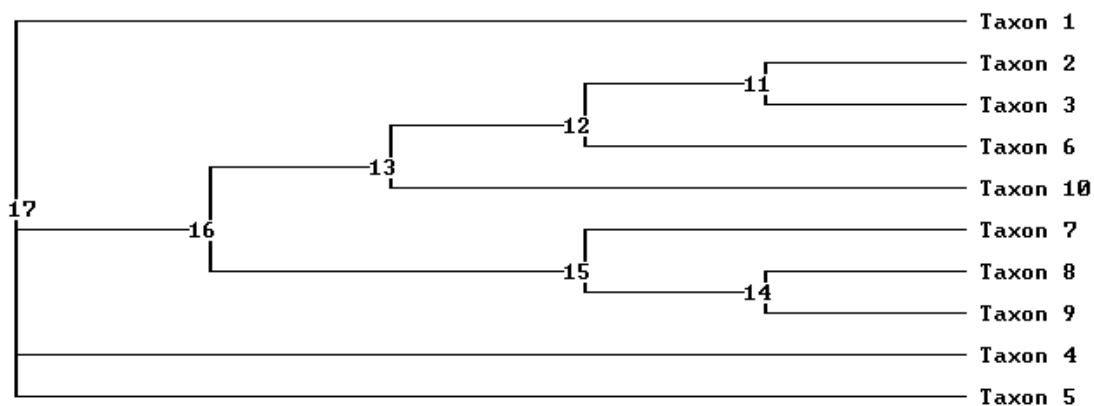


Figura 5: Cladograma da matriz 2 de 10 táxons/100 caracteres 25%.

Matriz 2 de 10 táxons/100 caracteres 100% (Figura 6)

Grupo-irmão: os táxons 8 e 9 estão em um grupo monofilético maior.

Quantidade de sinapomorfias no ramo 13>12 = 10 ambíguas e 8 não ambíguas.

Quantidade de caracteres excluídos no táxon 8 = 0

Quantidade de caracteres excluídos no táxon 9 = 0

Tree length = 281
 Consistency index (CI) = 0.3559
 Homoplasy index (HI) = 0.6441
 CI excluding uninformative characters = 0.3513
 HI excluding uninformative characters = 0.6487
 Retention index (RI) = 0.3489
 Rescaled consistency index (RC) = 0.1242

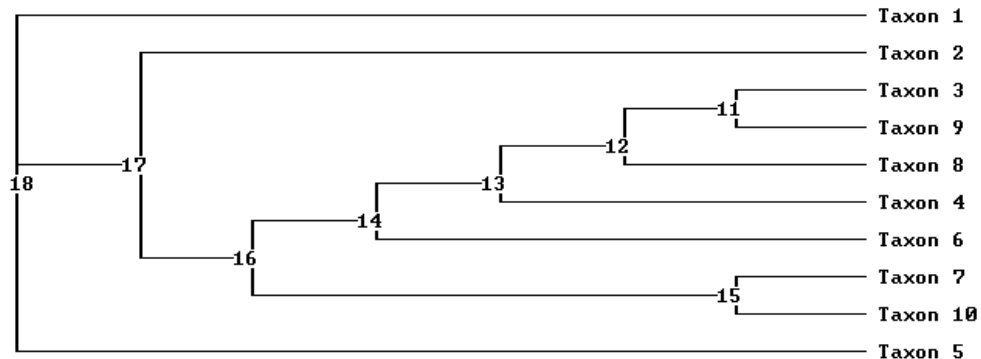


Figura 6: Cladograma da matriz 2 de 10 táxons/100 caracteres 100%.

Houve uma mudança na topologia da árvore, na qual a árvore de 100% agrupa os táxons 3 e 9 como grupos-irmãos e o taxón 8 faz parte do grupo monofilético que une os três (3,9,8). Na árvore de 25% temos o grupo-irmão 8 e 9 unidos ao táxon 7 formando outro grupo monofilético. Neste caso o número de sinapomorfias ambíguas é maior na árvore de 100%, o que pode caracterizar que as sinapomorfias que foram excluídas na árvore de 25% eram ambíguas. Observa-se no gráfico da matriz 2 (em anexo), entretanto, que o número de MPTs praticamente não aumenta com a exclusão dos caracteres. Por outro lado, há uma diminuição do número global de NSA e NSNA.

Matriz 6 de 10 táxons/100 caracteres 25% (Figura 7)

Grupo-irmão: 3 e 4

Quantidade de sinapomorfias no ramo 12>11= 1 ambígua e 2 não ambíguas.

Quantidade de caracteres excluídos no táxon 3 = 80 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 4 = 76 caracteres excluídos.

Tree length = 57
 Consistency index (CI) = 0.9123
 Homoplasy index (HI) = 0.0877
 CI excluding uninformative characters = 0.7059
 HI excluding uninformative characters = 0.2941
 Retention index (RI) = 0.6154
 Rescaled consistency index (RC) = 0.5614

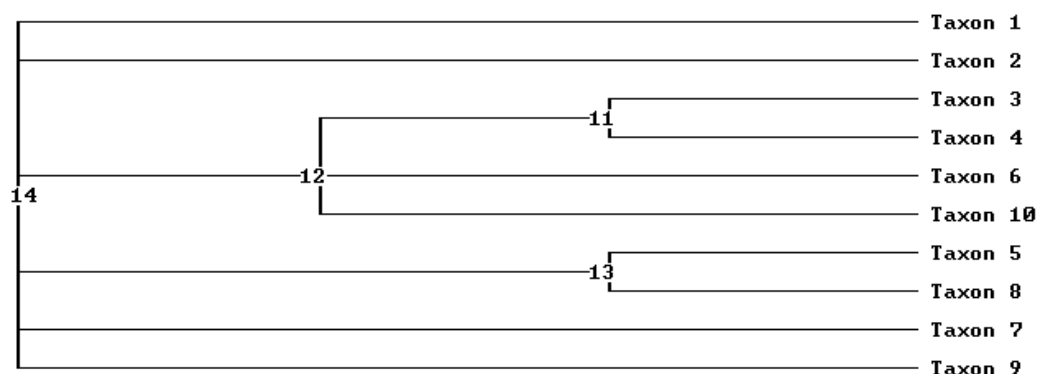


Figura 7: Cladograma da matriz 6 de 10 táxons/100 caracteres 25%.

Matriz 6 de 10 táxons/100 caracteres 100% (Figura 8)

Grupo-irmão 3 e 4 não estão juntos

Quantidade de sinapomorfias no ramo 17>16 = 5 ambíguas e 4 não ambíguas.

Quantidade de caracteres excluídos no táxon 3 = 0

Quantidade de caracteres excluídos no táxon 4 = 0

Tree length = 286
Consistency index (CI) = 0.3497
Homoplasy index (HI) = 0.6503
CI excluding uninformative characters = 0.3451
HI excluding uninformative characters = 0.6549
Retention index (RI) = 0.3357
Rescaled consistency index (RC) = 0.1174

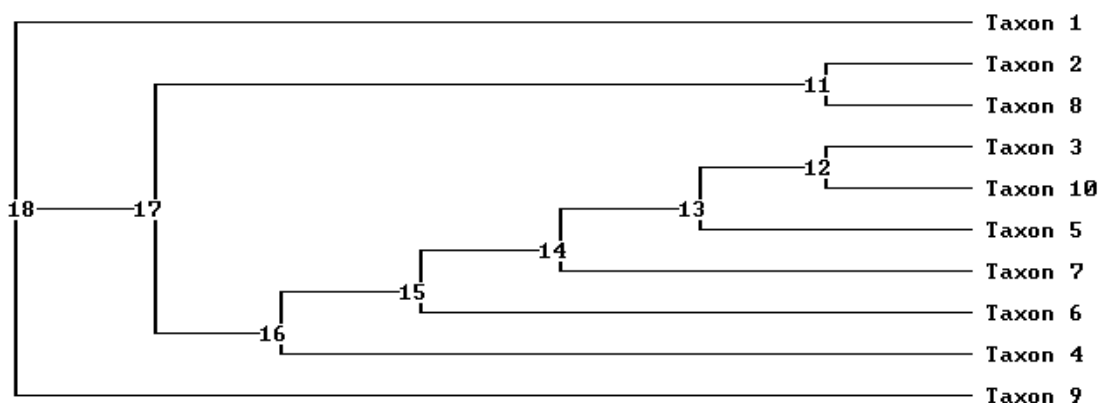


Figura 8: Cladograma da matriz 6 de 10 táxons/100 caracteres 100%.

Podemos observar que houve uma mudança na topologia da árvore, o grupo monofilético (3 e 4) que aparece na árvore da matriz de 25%, ocorre na árvore da matriz de 100% em um novo grupo monofilético que agrupa os táxons 3, 10, 5, 7, 6, 4. O número de sinapomorfias ambíguas que agrupa esses táxons é maior do que o número de sinapomorfias não ambíguas na árvore de 100%, se comparado com a árvore de 25%. Podemos inferir também que o número de caracteres excluídos das árvores de 25% eram sinapomorfias ambíguas, pois o grupo monofilético (3 e 4) permanece na árvore de 25%, e o número de sinapomorfias não ambíguas no nó que os une é maior que o número de sinapomorfias ambíguas, apesar dele também estar agrupado aos táxons 6 e 10. Observa-se no gráfico da matriz 6 (em anexo) que há um ligeiro aumento do número de MPTs na matriz de 25% e uma diminuição gradual do valor global de NSA e NSNA.

Matriz 8 de 10 táxons/100 caracteres 25% (Figura 9)

Grupo-irmão: 5 e 10

Quantidade de sinapomorfias no ramo 15>14 = 1 ambígua e 1 não ambígua

Quantidade de caracteres excluídos no táxon 5 = 76 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 10 = 81 caracteres excluídos.

Tree length = 54
 Consistency index (CI) = 0.9815
 Homoplasy index (HI) = 0.0185
 CI excluding uninformative characters = 0.9000
 HI excluding uninformative characters = 0.1000
 Retention index (RI) = 0.8889
 Rescaled consistency index (RC) = 0.8724

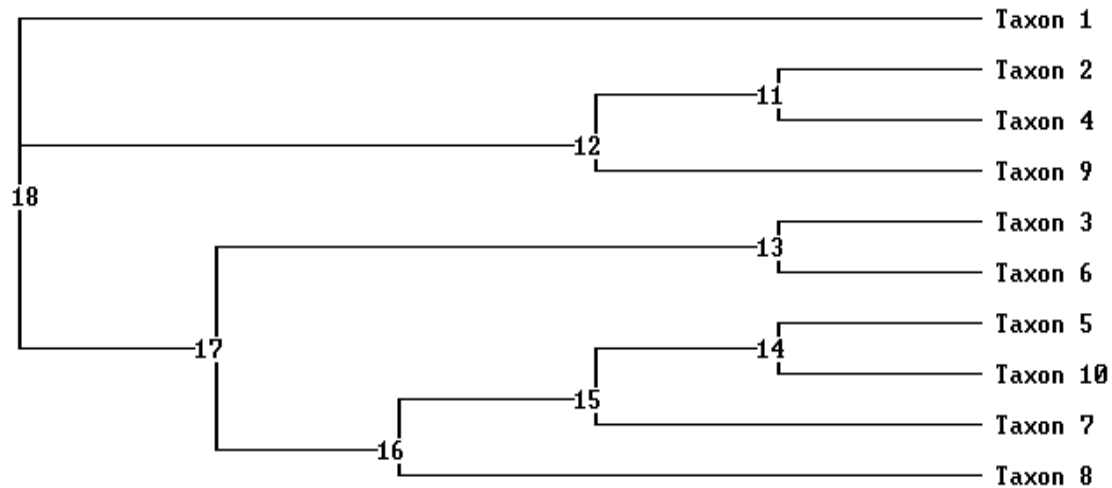


Figura 9: Cladograma da matriz 8 de 10 táxons/100 caracteres 25%.

Matriz 8 de 10 táxons/100 caracteres 100% (Figura 10)

Grupo-irmão 5 e 10 estão em grupos monofiléticos diferentes.

Quantidade de sinapomorfias no ramo 18>17 = 5 ambíguas e 9 não ambíguas.

Quantidade de caracteres excluídos no táxon = 0

Quantidade de caracteres excluídos no táxon = 0

Tree length = 278
 Consistency index (CI) = 0.3597
 Homoplasy index (HI) = 0.6403
 CI excluding uninformative characters = 0.3551
 HI excluding uninformative characters = 0.6449
 Retention index (RI) = 0.3333
 Rescaled consistency index (RC) = 0.1199

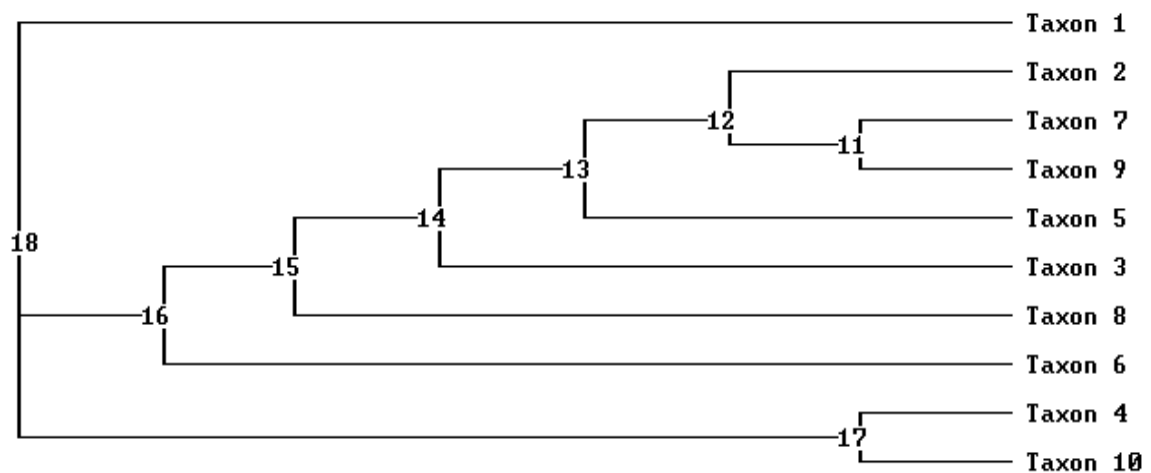


Figura 10: Cladograma da matriz 8 de 10 táxons/100 caracteres 100%.

Mais uma vez ocorre uma mudança na topologia da árvore, onde o grupo-irmão 5 e 10 não está presente na árvore original. Os dois estão em dois grupos monofiléticos distintos que fazem parte de uma politomia basal. Na árvore de 25% os táxons 5 e 10 são grupos-irmãos e estão agrupados com o táxon 7 em um outro grupo monofilético. O número de sinapomorfias não ambíguas é maior do que as ambíguas na árvore de 100%, o que pode implicar que um maior número de sinapomorfias não ambíguas foi excluído na árvore de 25%. Observa-se no gráfico da matriz 8 (em anexo) que o número de MPTs praticamente não muda e uma diminuição gradual do valor global de NSA e NSNA.

Matriz 10 de 10 táxons/100 caracteres 25% (Figura 11)

Grupo-irmão: 6 e 10

Quantidade de sinapomorfias no ramo 12 > 11 = 6 ambíguas e 2 não ambíguas

Quantidade de caracteres excluídos no táxon 6 = 77 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 10 = 68 caracteres excluídos.

Tree length = 58
 Consistency index (CI) = 0.9310
 Homoplasy index (HI) = 0.0690
 CI excluding uninformative characters = 0.8095
 HI excluding uninformative characters = 0.1905
 Retention index (RI) = 0.7647
 Rescaled consistency index (RC) = 0.7120

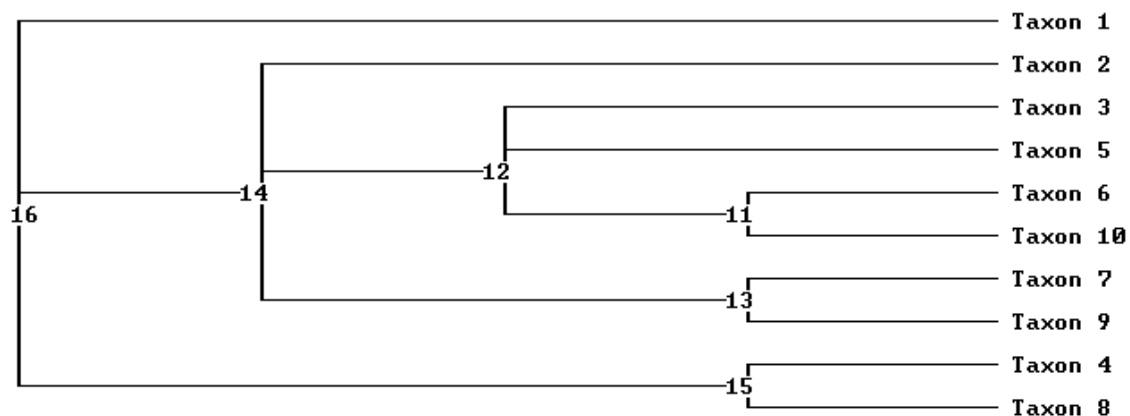


Figura 11: Cladograma da matriz 10 de 10 táxons/100 caracteres 25%.

Matriz 10 de 10 táxons/100 caracteres 100% (Figura 12)

Grupo-irmão: 6 e 10 estão em um grupo monofilético mais amplo

Quantidade de sinapomorfias no ramo 18 > 17 = 10 ambíguas e 7 não ambíguas

Quantidade de caracteres excluídos no táxon 6 = 0

Quantidade de caracteres excluídos no táxon 10 = 0

Tree length = 272
 Consistency index (CI) = 0.3676
 Homoplasy index (HI) = 0.6324
 CI excluding uninformative characters = 0.3485
 HI excluding uninformative characters = 0.6515
 Retention index (RI) = 0.3385
 Rescaled consistency index (RC) = 0.1244

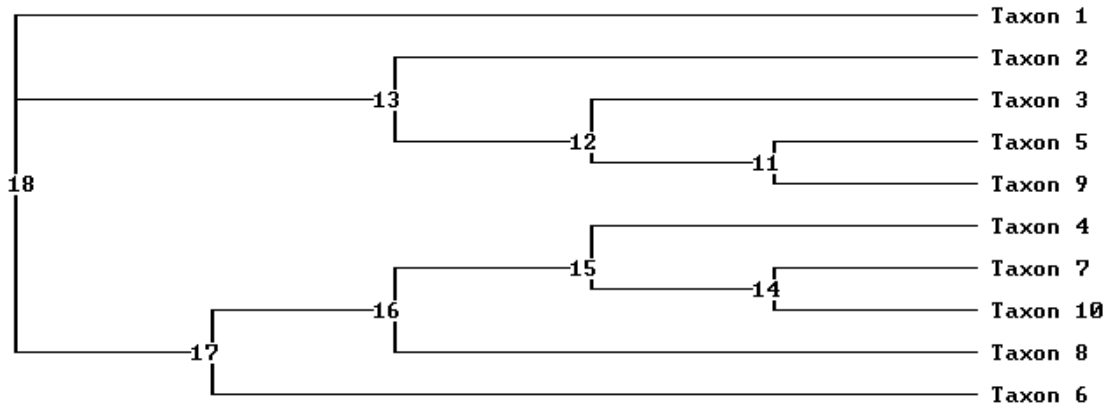


Figura 12: Cladograma da matriz 10 de 10 táxons/100 caracteres 100%.

Podemos verificar que novamente houve uma mudança na topologia da árvore, o grupo-irmão 6 e 10 formando um novo grupo monofilético com os táxons 3 e 5 que ocorre na árvore de 25%, não se repete na árvore de 100%, nesse caso eles formam um grupo monofilético com outros táxons: 4, 7 e 10, 8, 6. Nos dois casos o número de sinapomorfias ambíguas que os une é maior do que as sinapomorfias não ambíguas. Há uma possibilidade de as sinapomorfias excluídas da árvore de 25% serem sinapomorfias ambíguas. Observa-se no gráfico da matriz 10 (em anexo) que há um ligeiro aumento do número de MPTs na matriz de 50% e uma diminuição gradual do valor global de NSNA e um ligeiro aumento, seguido de diminuição do número global de NSA na matriz de 75%.

Matriz 17 de 10 táxons/100 caracteres 25% (Figura 13)

Grupo-irmão: 9 e 10

Quantidade de sinapomorfias no ramo 12>11 = 1 ambígua e 1 não ambígua

Quantidade de caracteres excluídos no táxon 9 = 72 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 10 = 80 caracteres excluídos.

Tree length = 60
 Consistency index (CI) = 1.0000
 Homoplasy index (HI) = 0.0000
 CI excluding uninformative characters = 1.0000
 HI excluding uninformative characters = 0.0000
 Retention index (RI) = 1.0000
 Rescaled consistency index (RC) = 1.0000

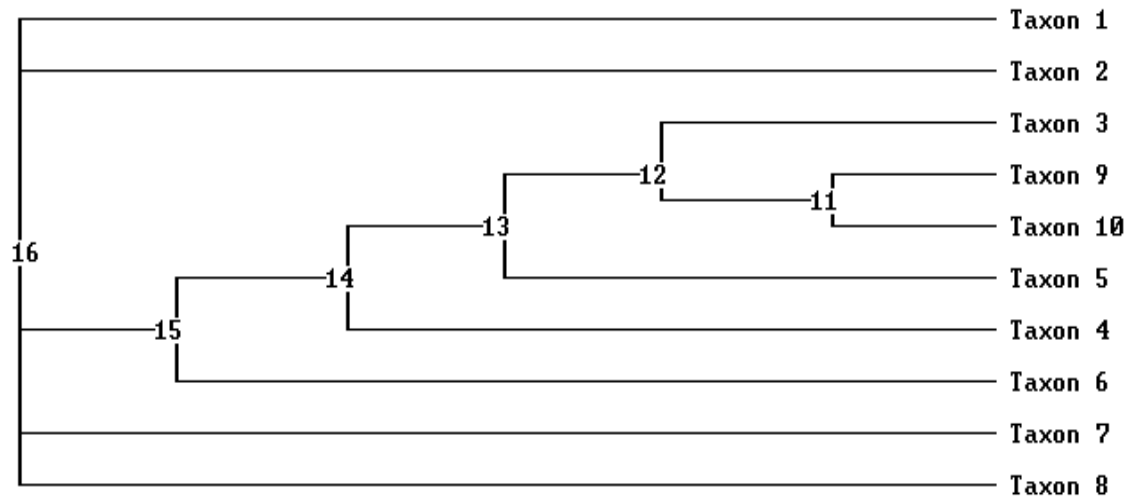


Figura 13: Cladograma da matriz 17 de 10 táxons/100 caracteres 25%.

Matriz 17 de 10 táxons/100 caracteres 100% (Figura 14)

Grupo-irmão: 9 e 10, estão associados a outros táxons e um grupo monofilético mais amplo.

Quantidade de sinapomorfias no ramo 18 > 17 = 5 ambíguas e 10 não ambíguas.

Quantidade de caracteres excluídos no táxon 9 = 0

Quantidade de caracteres excluídos no táxon 10 = 0

Tree length = 286
 Consistency index (CI) = 0.3462
 Homoplasy index (HI) = 0.6538
 CI excluding uninformative characters = 0.3439
 HI excluding uninformative characters = 0.6561
 Retention index (RI) = 0.3345
 Rescaled consistency index (RC) = 0.1158

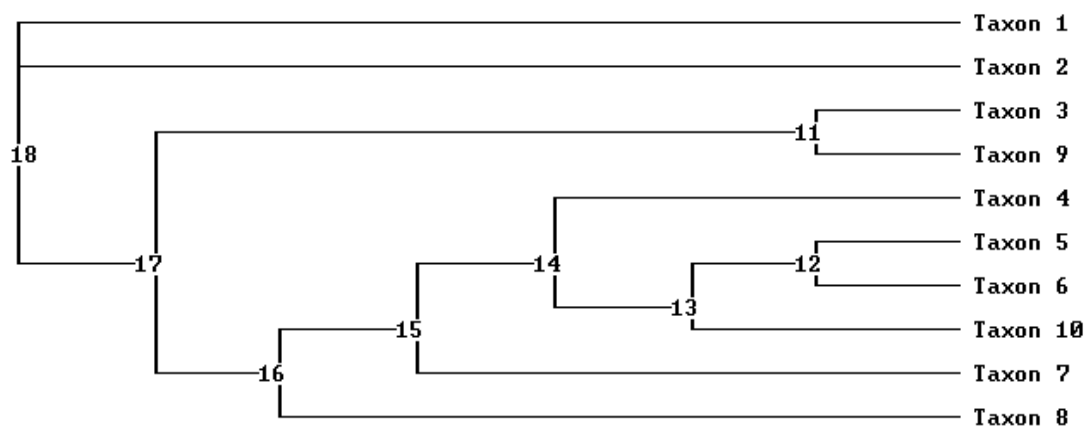


Figura 14: Cladograma da matriz 17 de 10 táxons/100 caracteres 100%.

A árvore de 100% apresenta uma topologia modificada com relação a árvore de 25%, nela identificamos um grupo monofilético mais amplo que inclui os táxons: 3 e 9, 4, 5 e 6, 10, 7, 8, diferentemente da árvore de 25% onde temos o grupo-irmão 9 e 10 formando outro grupo monofilético com o táxon 3. O número de sinapomorfias não ambíguas é maior que o número de sinapomorfias ambíguas na árvore 100%. Nesse caso, um número maior de sinapomorfias não ambíguas pode ter sido excluído na árvore de 25%. Observa-se no gráfico da matriz 17 (em anexo) que o número de MPTs praticamente não muda e que há uma diminuição gradual do valor global de NSNA e que o valor global de NSA diminui apenas na matriz de 25%.

Para as matrizes de 20 táxons e 100 caracteres obtivemos o seguinte resultado:

Matriz 2 de 20 táxons/100 caracteres 25% (Figura 15)

Grupo-irmão: grupo 18>19

Quantidade de sinapomorfias no ramo 27>26 = 5 ambíguas e 3 não ambíguas.

Quantidade de caracteres excluídos no táxon 18 = 75 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 19 = 69 caracteres excluídos.

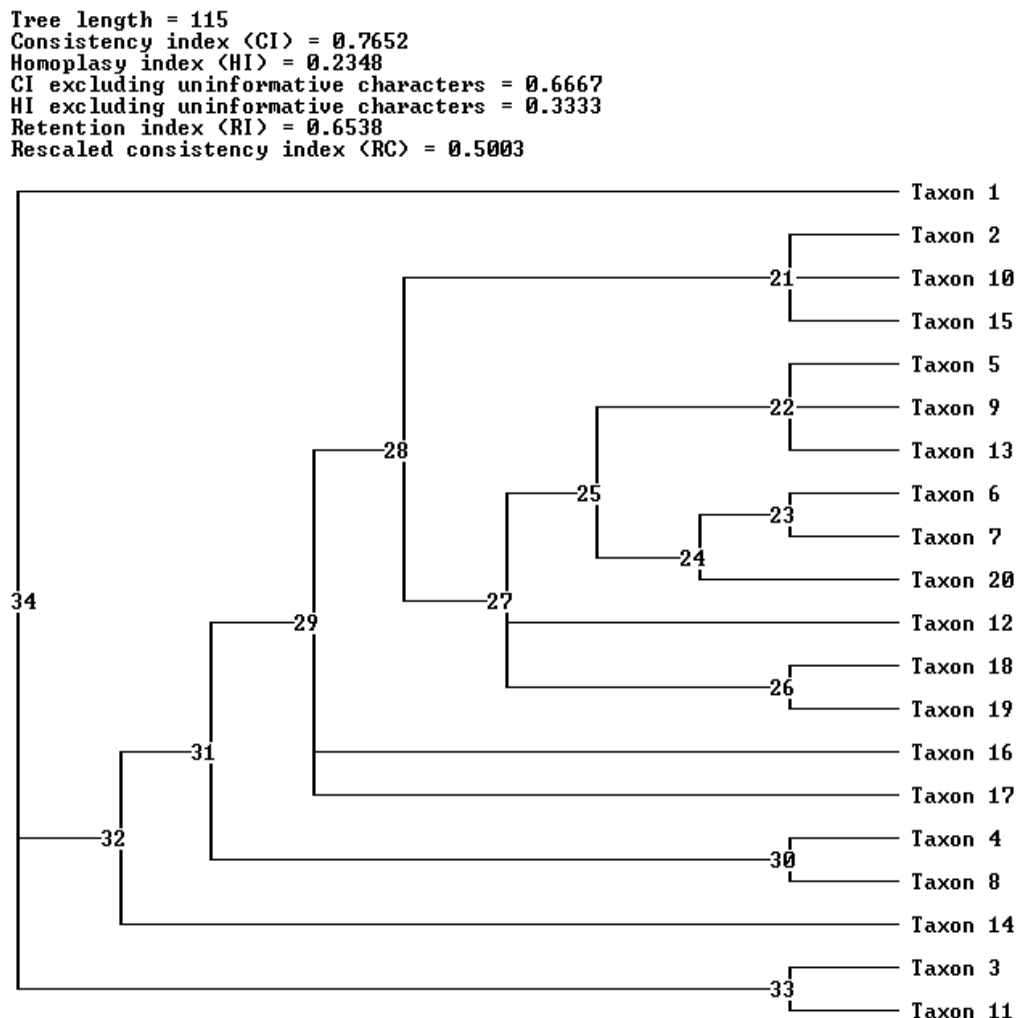


Figura 15: Cladograma da matriz 2 de 20 táxons/100 caracteres 25%.

Matriz 2 de 20 táxons/100 caracteres 100% (Figura 16)

Grupo-irmão: 18 e 19, os táxons apresentam configuração completamente distinta.

Quantidade de sinapomorfias no ramo 38>37 = 4 ambíguas e 12 não ambíguas.

Quantidade de caracteres excluídos no táxon 18 = 0

Quantidade de caracteres excluídos no táxon 19 = 0

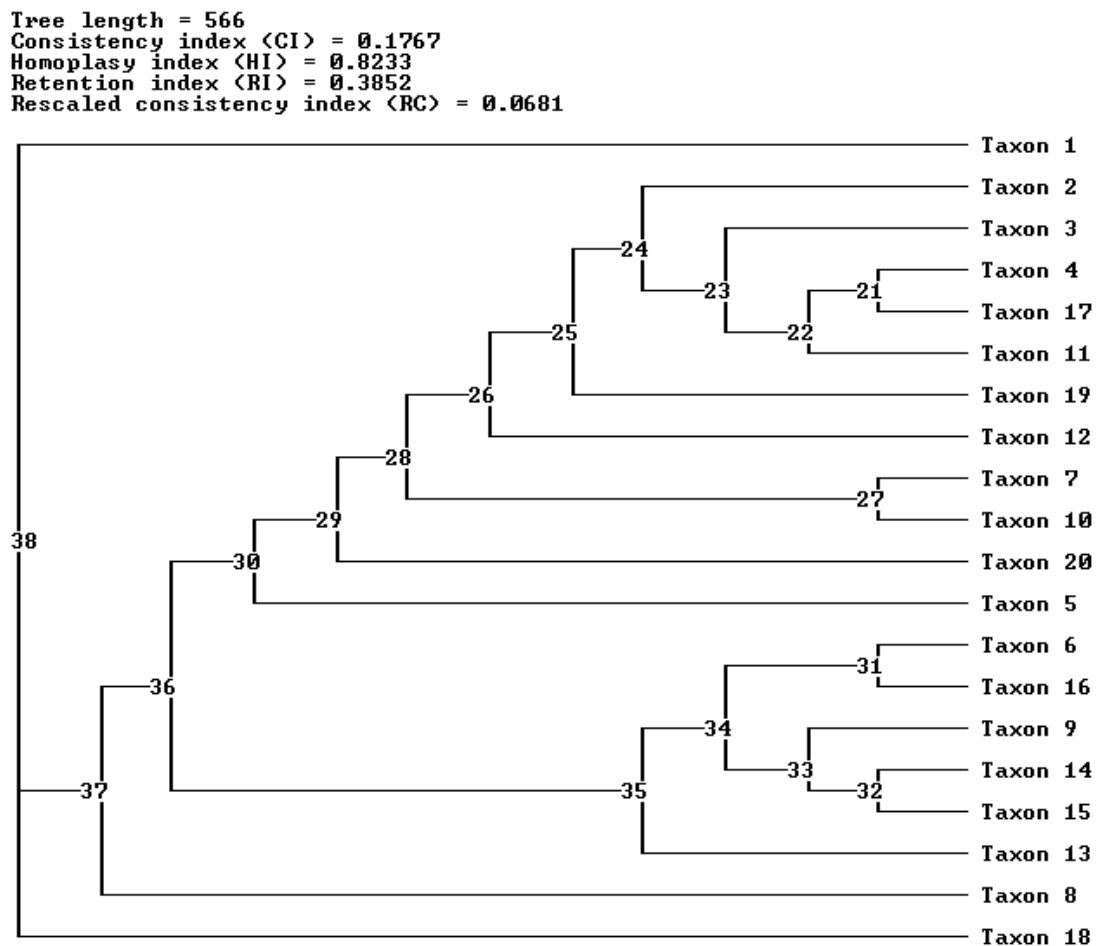


Figura 16: Cladograma da matriz 2 de 20 táxons/100 caracteres 100%.

A topologia da árvore de 100% foi modificada em relação à árvore de 25%, na qual não encontramos o grupo-irmão 18 e 19, agora eles fazem parte de uma politomia basal que inclui praticamente todos os outros táxons da árvore. Diferentemente, ocorre na árvore de 25%, onde o grupo-irmão está em outro grupo monofilético que inclui também os táxons 5, 9 e 13, 6 e 7, 20, 12. A árvore de 100% apresenta um número maior de sinapomorfias não ambíguas, já a árvore de 25% apresenta um número maior de sinapomorfias ambíguas, isso pode representar que um grande número de sinapomorfias não ambíguas foram excluídas da árvore de 25%. Observa-se no gráfico da matriz 2 (em anexo) que há um aumento do número de MPTs a partir da matriz de 50% e uma diminuição gradual do valor global de NSA e NSNA.

Matriz 5 de 20 táxons/100 caracteres 25% (Figura 17)

Grupo-irmão: 2 e 16

Quantidade de sinapomorfias no ramo 24>21 = 2 ambíguas e 1 não ambígua.
 Quantidade de caracteres excluídos no táxon 2 = 72 caracteres excluídos.
 Quantidade de caracteres excluídos no táxon 16 = 82 caracteres excluídos.

Tree length = 119
 Consistency index (CI) = 0.7143
 Homoplasy index (HI) = 0.2857
 CI excluding uninformative characters = 0.6383
 HI excluding uninformative characters = 0.3617
 Retention index (RI) = 0.6092
 Rescaled consistency index (RC) = 0.4351

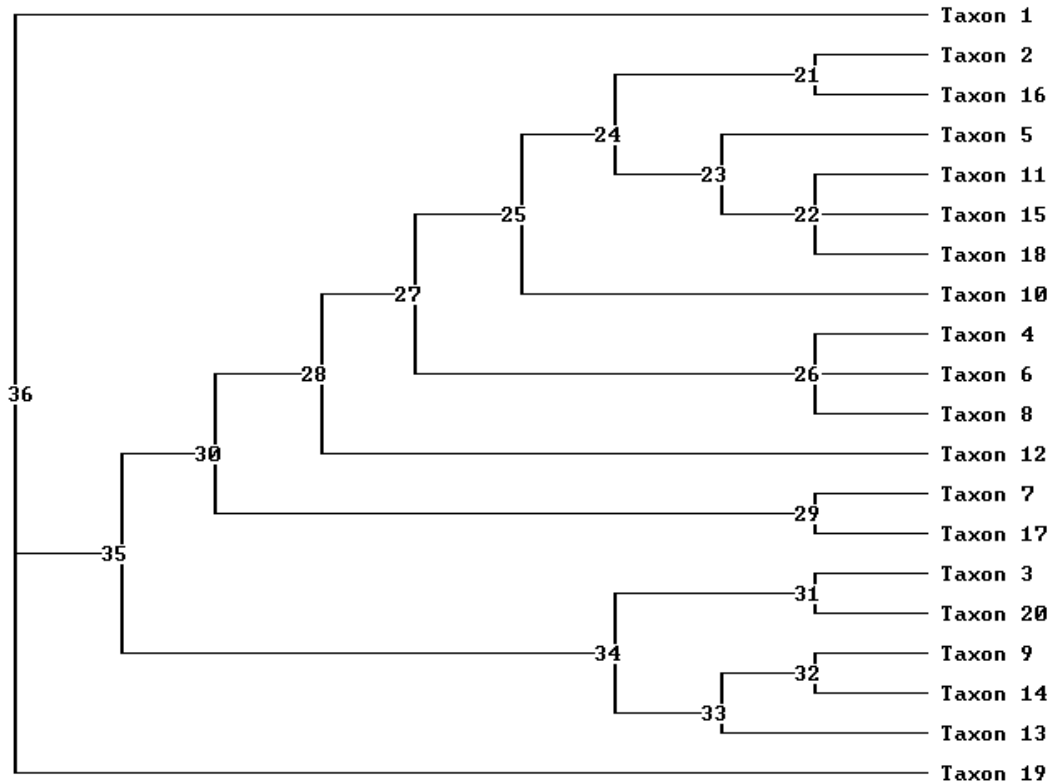


Figura 17: Cladograma da matriz 5 de 20 táxons/100 caracteres 25%.

Matriz 5 de 20 táxons/100 caracteres 100% (Figura 18)

Grupo-irmão: 2 e 16, fazem parte de um grupo monofilético com o táxon 7, mas não aparecem como grupos-irmãos

Quantidade de sinapomorfias no ramo 23>22 = 12 ambíguas e 6 não ambíguas.

Quantidade de caracteres excluídos no táxon 2 = 0

Quantidade de caracteres excluídos no táxon 16 = 0

Tree length = 561
 Consistency index (CI) = 0.1783
 Homoplasy index (HI) = 0.8217
 Retention index (RI) = 0.3719
 Rescaled consistency index (RC) = 0.0663

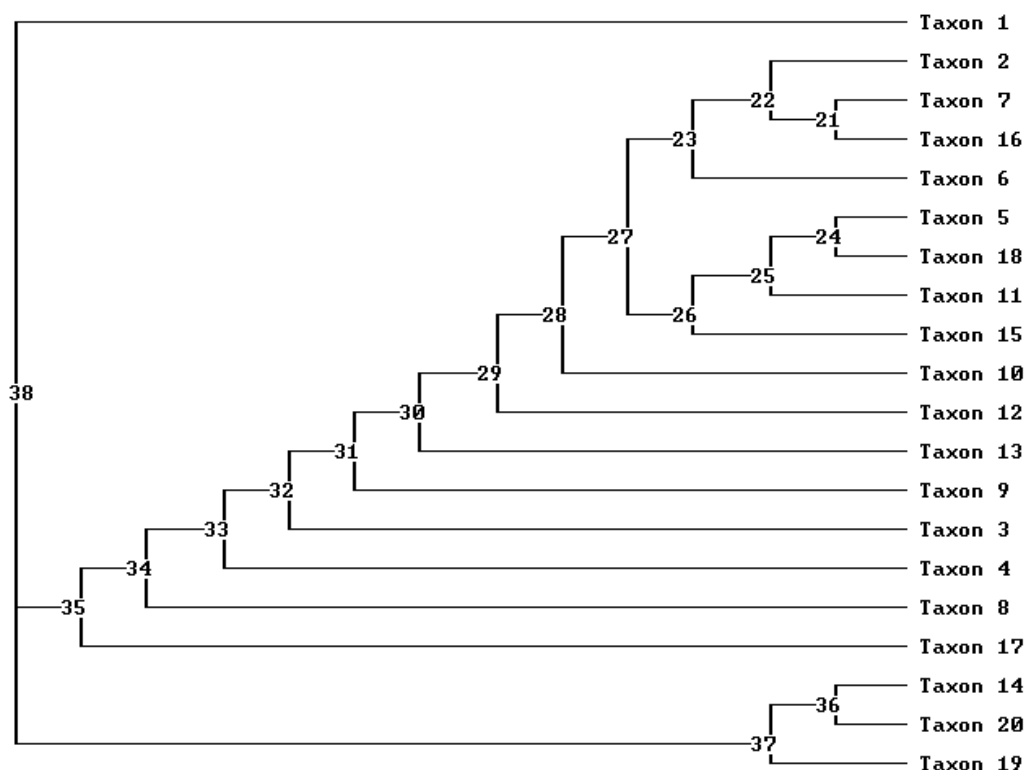


Figura 18: Cladograma da matriz 5 de 20 táxons/100 caracteres 100%.

Houve uma mudança na topologia da árvore de 100%, onde o grupo-irmão 2 e 16 (na árvore de 25%) não ocorre, mas esses táxons formam um grupo monofilético com o táxon 7, na árvore de 25% o grupo-irmão 2 e 16 estão em um outro grupo monofilético que inclui os táxons 5, 11, 15 e 18. O número de sinapomorfias não ambíguas é menor nos dois casos, isso pode ser devido ao número maior de sinapomorfias não ambíguas excluídas dessas árvores. Observa-se no gráfico da matriz 5 (em anexo) que o número de MPTs praticamente não muda até a matriz de 75%, iniciando um aumento na matriz de 50% e se tornando muito acentuado na matriz de 25%. Por outro lado, há uma diminuição gradual do valor global de NSA e NSNA.

Matriz 7 de 20 táxons/100 caracteres 25% (Figura 19)

Grupo-irmão: 13 e 19

Quantidade de sinapomorfias no ramo 29>28 = 3 ambíguas e 1 não ambígua.

Quantidade de caracteres excluídos no táxon 13 = 81 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 19 = 73 caracteres excluídos.

Tree length = 111
 Consistency index (CI) = 0.7658
 Homoplasy index (HI) = 0.2342
 CI excluding uninformative characters = 0.6389
 HI excluding uninformative characters = 0.3611
 Retention index (RI) = 0.6389
 Rescaled consistency index (RC) = 0.4892

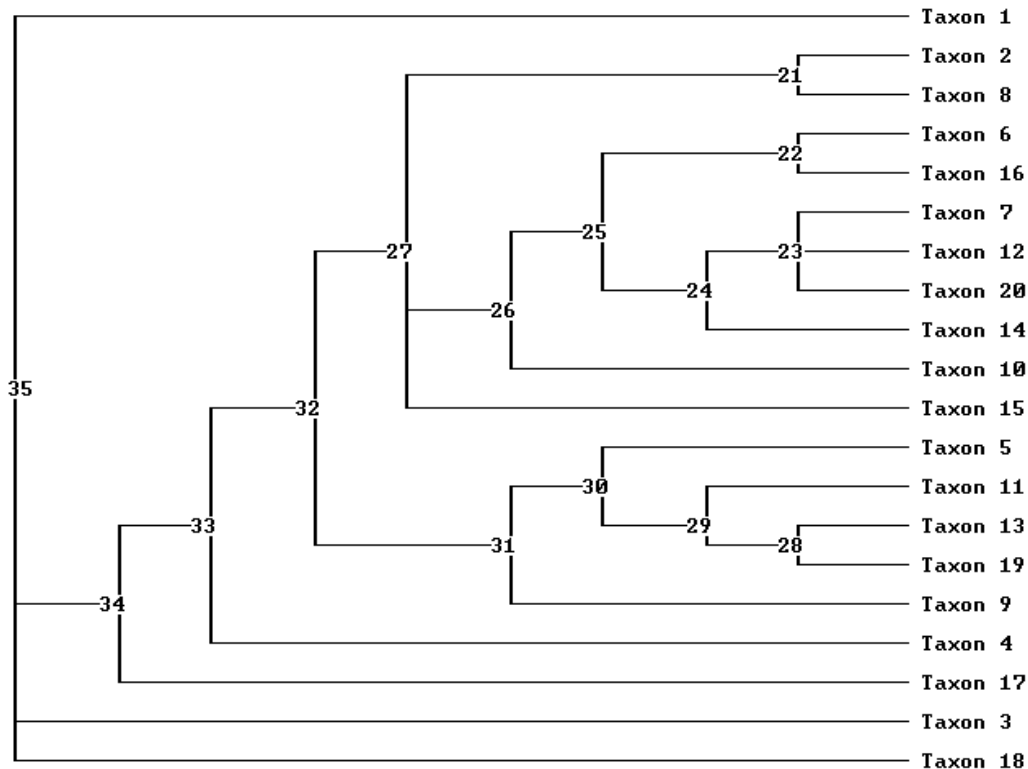


Figura 19: Cladograma da matriz 7 de 20 táxons/100 caracteres 25%.

Matriz 7 de 20 táxons/100 caracteres 100% (Figura 20)

Grupo-irmão: 13 e 19, estão em grupos monofiléticos distintos dentro de um grupo monofilético mais amplo.

Quantidade de sinapomorfias no ramo 34>33 = 7 ambíguas e 9 não ambíguas.

Quantidade de caracteres excluídos no táxon 13 = 0

Quantidade de caracteres excluídos no táxon 19 = 0

Tree length = 569
 Consistency index (CI) = 0.1757
 Homoplasy index (HI) = 0.8243
 Retention index (RI) = 0.3549
 Rescaled consistency index (RC) = 0.0624

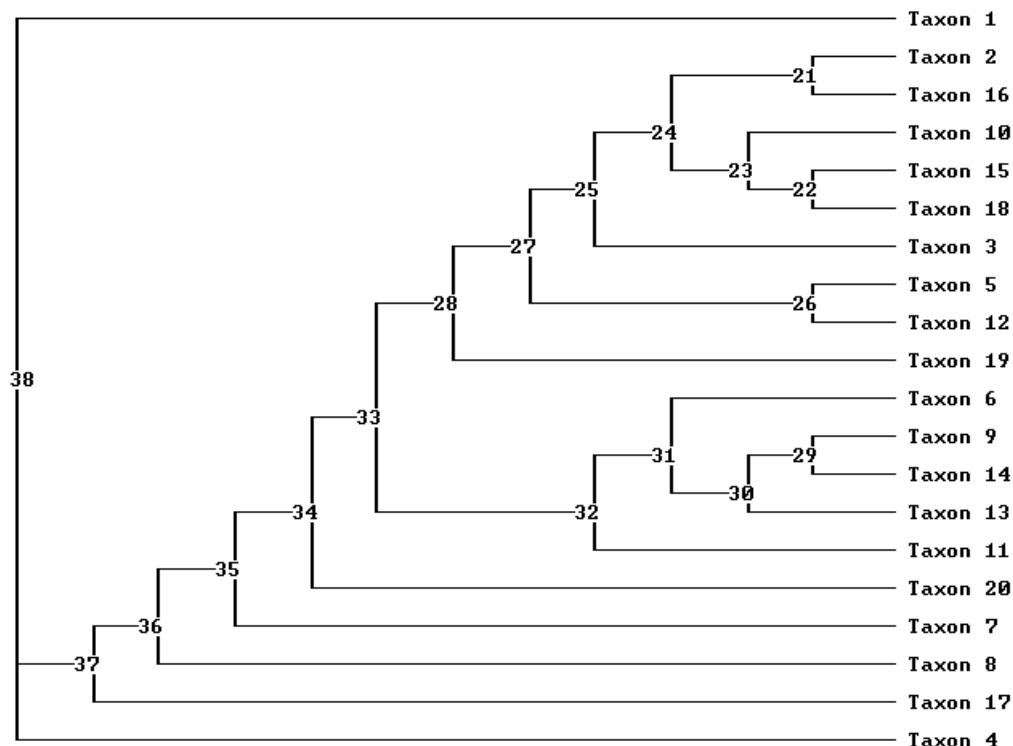


Figura 20: Cladograma da matriz 7 de 20 táxons/100 caracteres 100%.

Novamente houve uma mudança na topologia da árvore de 100%, na qual não podemos identificar os táxons 13 e 19 como grupos-irmãos como ocorre na árvore de 25%. Eles formam um grupo monofilético que inclui os táxons 6, 9, 11 e 14, e outro grupo monofilético com os táxons 2 e 16, 10, 15 e 18, 3, 5 e 12, respectivamente. Enquanto que na árvore de 25% formam um grupo monofilético com o táxon 11. O número de sinapomorfias não ambíguas é maior que o número de sinapomorfias ambíguas na árvore de 100%, já na árvore de 25% ocorre o contrário, pode-se inferir novamente que isso pode ser devido ao número maior de sinapomorfias não ambíguas que foram excluídas dessa árvore. Observa-se no gráfico da matriz 7 (em anexo) que há um ligeiro aumento do número de MPTs na matriz de 50%, caindo um pouco na matriz de 25%. Há também uma diminuição gradual do valor global de NSNA, enquanto que o valor de NSA diminui mais acentuadamente apenas na matriz de 25%.

Matriz 9 de 20 táxons/100 caracteres 25% (Figura 21)

Grupo-irmão: 9 e 19

Quantidade de sinapomorfias no ramo 26>25 = 3 ambíguas e 1 não ambíguas.

Quantidade de caracteres excluídos no táxon 9 = 79 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 19 = 76 caracteres excluídos.

Tree length = 116
 Consistency index (CI) = 0.7672
 Homoplasy index (HI) = 0.2328
 CI excluding uninformative characters = 0.6197
 HI excluding uninformative characters = 0.3803
 Retention index (RI) = 0.6197
 Rescaled consistency index (RC) = 0.4755

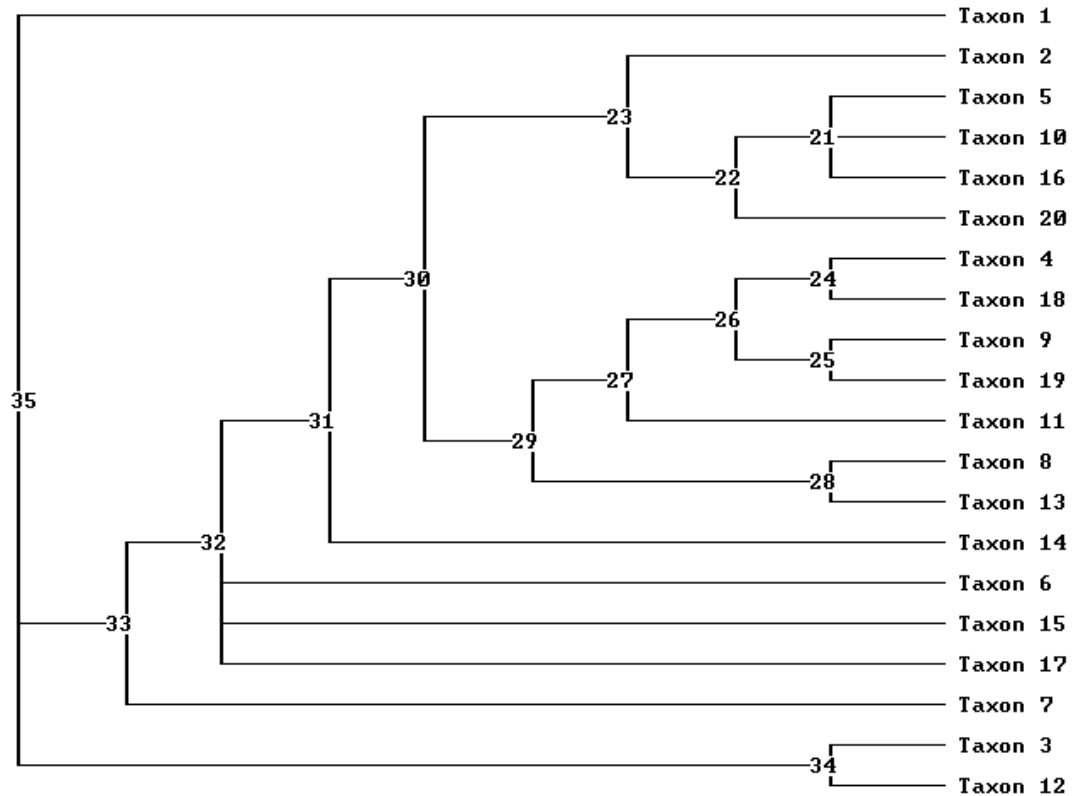


Figura 21: Cladograma da matriz 9 de 20 táxons/100 caracteres 25%.

Matriz 9 de 20 táxons/100 caracteres 100% (Figura 22)

Grupo-irmão: 9 e 19, estão em um grupo monofilético mais amplo

Quantidade de sinapomorfias no ramo 32>31 = 5 ambíguas e 5 não ambíguas.

Quantidade de caracteres excluídos no táxon 9 = 0

Quantidade de caracteres excluídos no táxon 19 = 0

Tree length = 560
 Consistency index <CI> = 0.1786
 Homoplasy index <HI> = 0.8214
 Retention index <RI> = 0.3699
 Rescaled consistency index <RC> = 0.0660

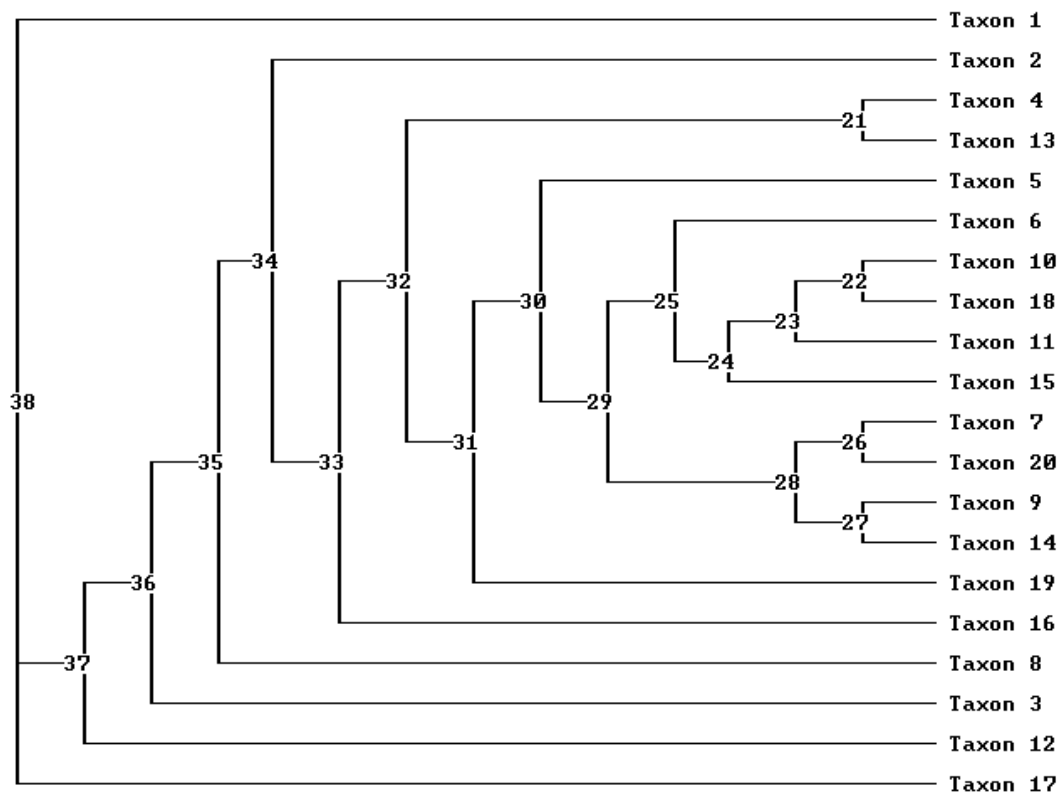


Figura 22: Cladograma da matriz 9 de 20 táxons/100 caracteres 100%.

A topologia da árvore foi novamente modificada, pois não identificamos na árvore 100% o grupo-irmão 9 e 19 presente na árvore 25%, na qual forma outro grupo monofilético com o grupo-irmão 4 e 18. Já na árvore de 100%, esses táxons estão incluídos em um grupo monofilético, não como grupos-irmãos, mas estão presentes nesse grupo que inclui também os táxons 9 e 14, 7 e 20, 15, 11, 10 e 18, 6, 5, 4 e 13. O número de sinapomorfias é igual para as ambíguas e não ambíguas na árvore de 100% e para a árvore de 25% o número de sinapomorfias ambíguas é maior. Podem ter sido excluídas mais sinapomorfias não ambíguas nessa árvore. Observa-se no gráfico da matriz 9 (em anexo) que há um aumento acentuado do número de MPTs na matriz de 25% e uma diminuição gradual do valor global de NSA e NSNA.

Matriz 11 de 20 táxons/100 caracteres 25% (Figura 23)

Grupo-irmão: 5 e 15

Quantidade de sinapomorfias no ramo 27>26 = 7 ambíguas e 1 não ambígua.

Quantidade de caracteres excluídos no táxon 5 = 82 caracteres excluídos.

Quantidade de caracteres excluídos no táxon 15 = 75 caracteres excluídos.

Tree length = 118
 Consistency index (CI) = 0.7288
 Homoplasy index (HI) = 0.2712
 CI excluding uninformative characters = 0.6190
 HI excluding uninformative characters = 0.3810
 Retention index (RI) = 0.5789
 Rescaled consistency index (RC) = 0.4219

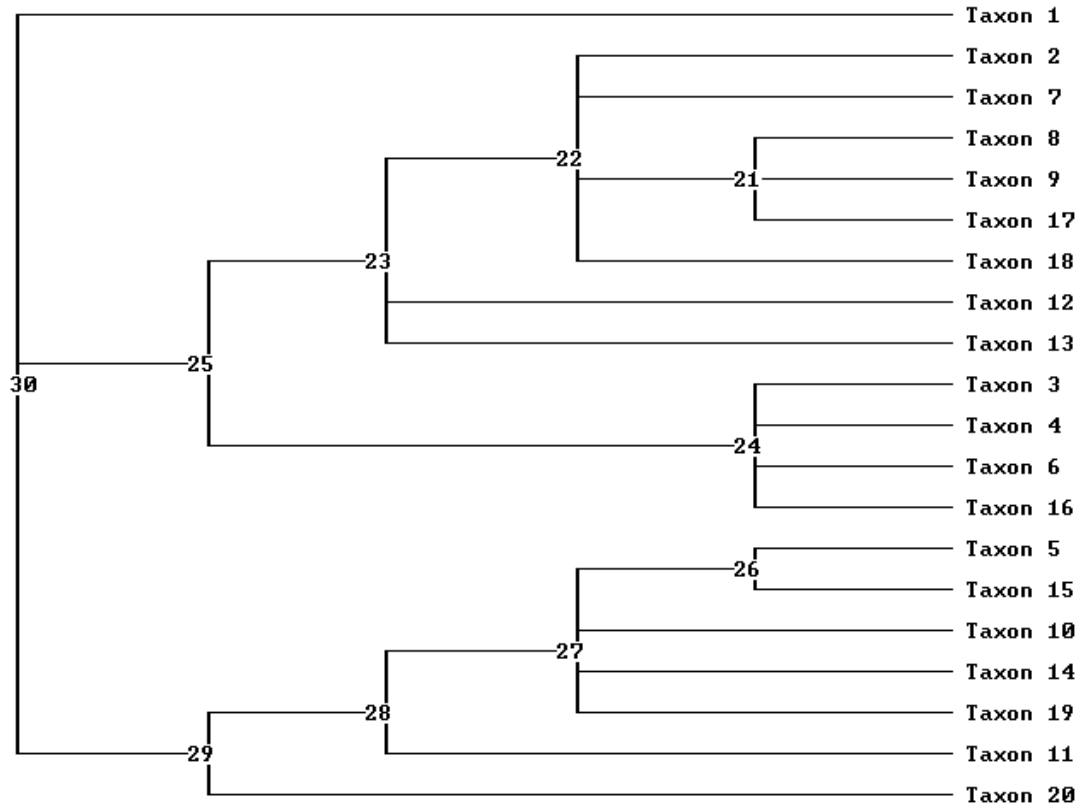


Figura 23: Cladograma da matriz 11 de 20 táxons/100 caracteres 25%.

Matriz 11 de 20 táxons/100 caracteres 100% (Figura 24)

Grupo-irmão: 5 e 15

Quantidade de sinapomorfias no ramo 38>37 = 3 ambíguas e 9 não ambíguas.

Quantidade de caracteres excluídos no táxon 5 = 0

Quantidade de caracteres excluídos no táxon 10 = 0

Tree length = 564
 Consistency index (CI) = 0.1773
 Homoplasy index (HI) = 0.8227
 Retention index (RI) = 0.3600
 Rescaled consistency index (RC) = 0.0638

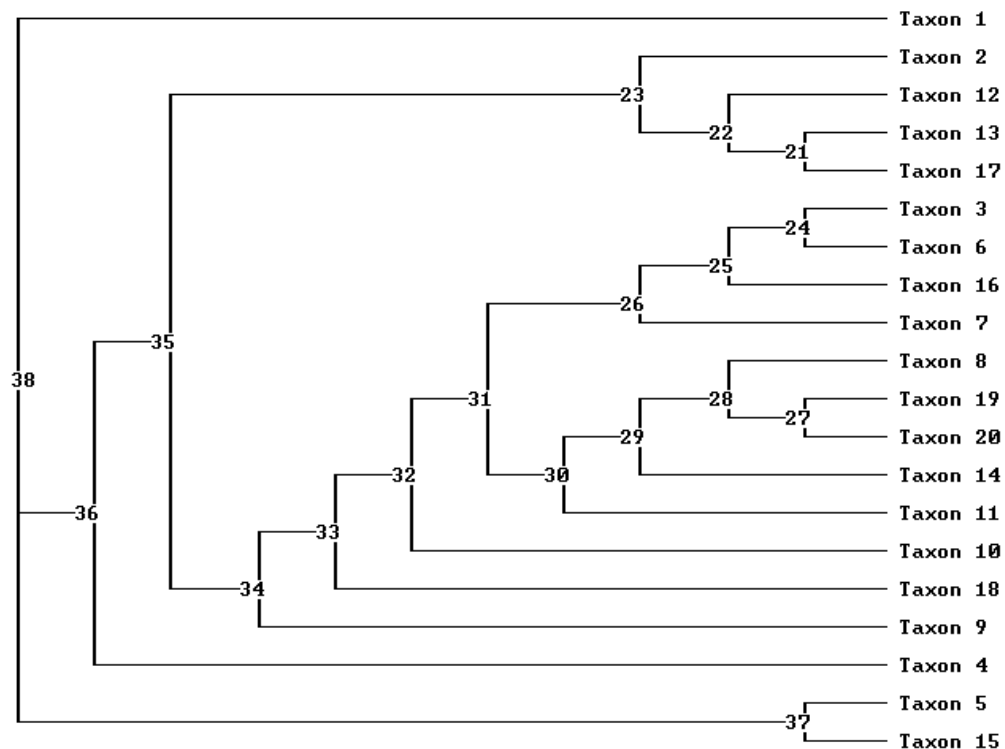


Figura 24: Cladograma da matriz 11 de 20 táxons/100 caracteres 100%.

Neste caso, o grupo-irmão 5 e 15 se manteve nas duas árvores, embora na árvore 25% ele faça parte de outro grupo monofilético com os táxons 10, 14, 19, formando uma politomia. Na árvore 100% esse grupo-irmão se manteve, mas faz parte de uma politomia basal que inclui praticamente todos os demais táxons. O número de sinapomorfias não ambíguas é maior na árvore 100%, e na árvore de 25% o número de sinapomorfias ambíguas é bem maior que o número de não ambíguas, isso pode ser devido às exclusões, nesta árvore, de caracteres que representavam sinapomorfias não ambíguas. Observa-se no gráfico da matriz 11 (em anexo) que há um acentuado aumento do número de MPTs na matriz de 25% e uma diminuição gradual do valor global de NSNA e um ligeiro aumento do valor de NSA na matriz de 75%, seguido de uma diminuição nas matrizes de 50% e 25%.

4. CONCLUSÃO

Com o trabalho desenvolvido foi possível inferir que em média o número de caracteres excluídos para um táxon nas matrizes de 25% varia de 70 – 80 % na exclusão aleatória. E com a mesma, aparentemente muitas sinapomorfias foram excluídas, o que ocasiona mudanças na posição dos táxons na árvore.

Em apenas um dos 10 casos analisados o grupo-irmão presente na árvore gerada na matriz de 100% se manteve na matriz de 25% (matriz 11 de 20 táxons). Nesse caso, observou-se que o número de sinapomorfias ambíguas aumentou. Podendo indicar que os caracteres excluídos se tratavam de plesiomorfias ou sinapomorfias não ambíguas que foram excluídas

em apenas um dos dois táxons. Nesse caso, passando de sinapomorfias não ambíguas para sinapomorfias ambíguas. Para se ter certeza absoluta disso, cada caráter em cada táxon dessas matrizes deveria ser checado manualmente antes e depois da exclusão para se averiguar se houve uma maior exclusão de plesiomorfias. Entretanto, devido ao tempo que essa tarefa requer, esse procedimento está fora do escopo desse trabalho. Ainda, mesmo que o grupo-irmão tenha se mantido nesse exemplo, destaca-se que a posição dele dentro da árvore foi modificada.

Com relação ao comportamento geral da árvore, pode-se concluir que a exclusão de caracteres mesmo que elevada (75% dos caracteres excluídos) não necessariamente diminui a resolução da árvore, ou seja, pode gerar um número baixo de MPTs em alguns casos. Além disso, e como um resultado inesperado, em todos os casos a exclusão de caracteres causou uma mudança na topologia das árvores quando comparada com a árvore original.

Assim, de acordo com os resultados obtidos, não podemos afirmar que existe um limite mínimo de *missing data* para a inclusão ou não de um táxon fóssil em uma análise filogenética. Entretanto, observa-se que geralmente táxons incompletos causam problemas de resolução na árvore.

5. PRÓXIMOS PASSOS

Sugere-se que sejam realizados estudos mais aprofundados e mais abrangentes envolvendo matrizes englobando táxons muito incompletos com táxons completos, na tentativa de melhor elucidar o comportamento dos táxons incompletos em análises envolvendo fósseis. Estudar o comportamento da topologia das árvores com a exclusão de informações da matriz, como podemos perceber nas árvores resultantes das matrizes de 100% e 25% observadas neste trabalho que, em todos os casos, a topologia foi alterada.

Modificar o programa GMMD para que seja possível excluir informações de um táxon ou caráter específico.

REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, D. S. **Fundamentos de Sistemática Filogenética**. São Paulo: Holos, 2002.

BENTON, M. J. 1990. The species of *Rhynchosaurus*, a rhynchosaur (Reptilia, Diapsida) from the middle Triassic of England. **Philosophical Transactions of the Royal Society of London** 328:213-306.

FRASER, N. C; BENTON, M. J. 1989. The Triassic reptiles *Brachyrhinodon* and *Polysphenodon* and the relationships of the sphenodontids. **Zoological Journal of the Linnean Society** 96:413-445.

GAO, K; NORELL, M. A. 1998. Taxonomic review of *Carusia* (Reptilia: Squamata) from the late Cretaceous of the Gobi Desert and phylogenetic relationships of anguimorph lizards. **American Museum Novitates** 3230:1-51.

GMMD. Gerador de matriz e *missing data*. Disponível em <http://www.sindicatodosbordes.com.br/gmmd/index.php>.

GRANDE, L; BEMIS, W. E. 1998. A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history. **Society of Vertebrate Paleontology** Memoir 4.

HARCOURT-BROWN, K.G; PEARSON, P. N; WILKINSON, M. The imbalance of paleontological trees. **Paleobiology**, Vol 27, No. 2, 2001, pp. 188-204.

KEARNEY, M; CLARK, J. M. Problems due to missing data in phylogenetic analyses including fossils: a critical review. **Journal of Vertebrate Paleontology**, vol. 23(2):263-274, 2003.

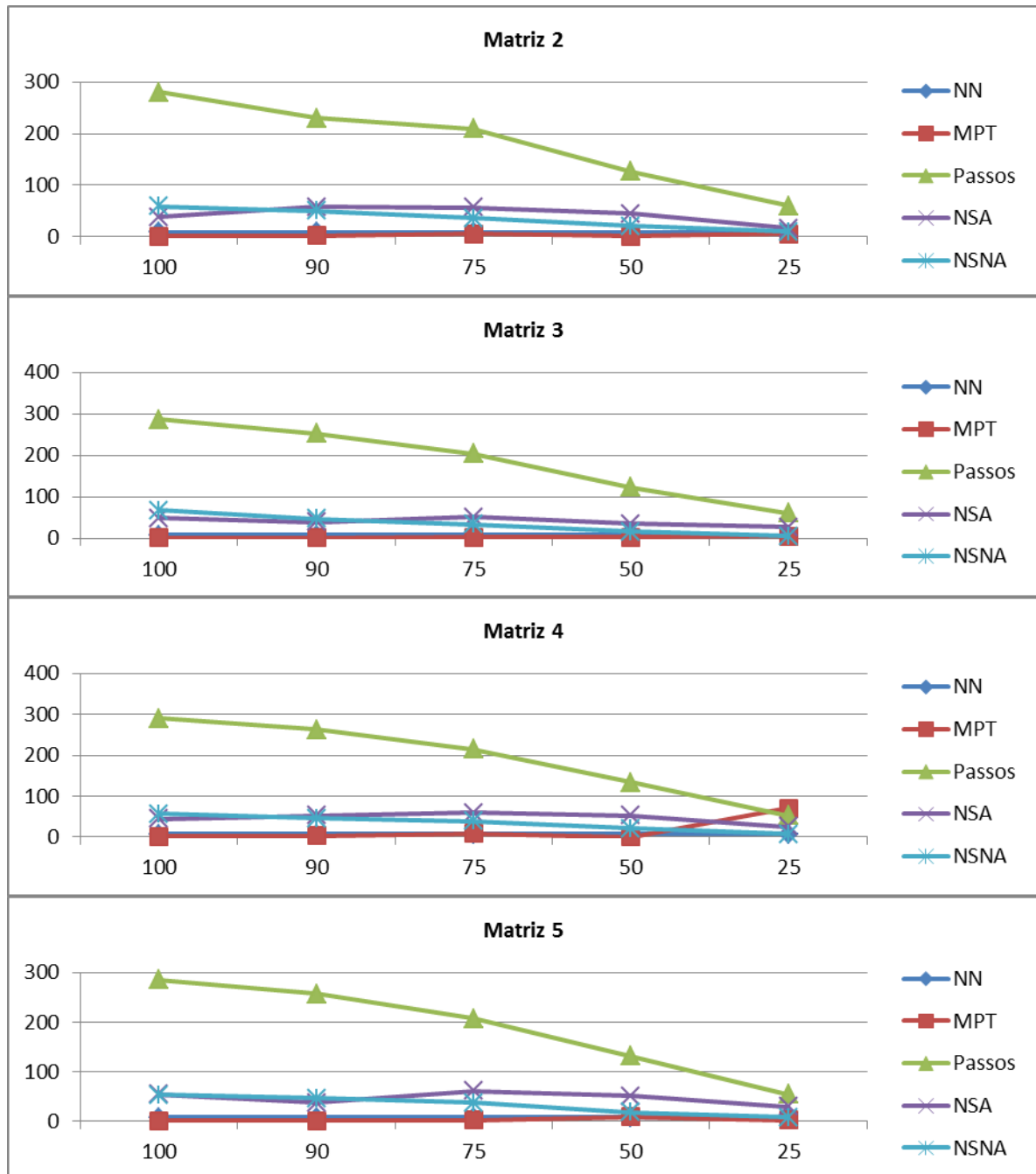
RIOS-NETO, A. M. **Paleontologia: conceitos e métodos**, cap. 11, volume 1/ editor, Ismar de Souza Carvalho. – 3ª ed. – Rio de Janeiro: Interciência, 2010.

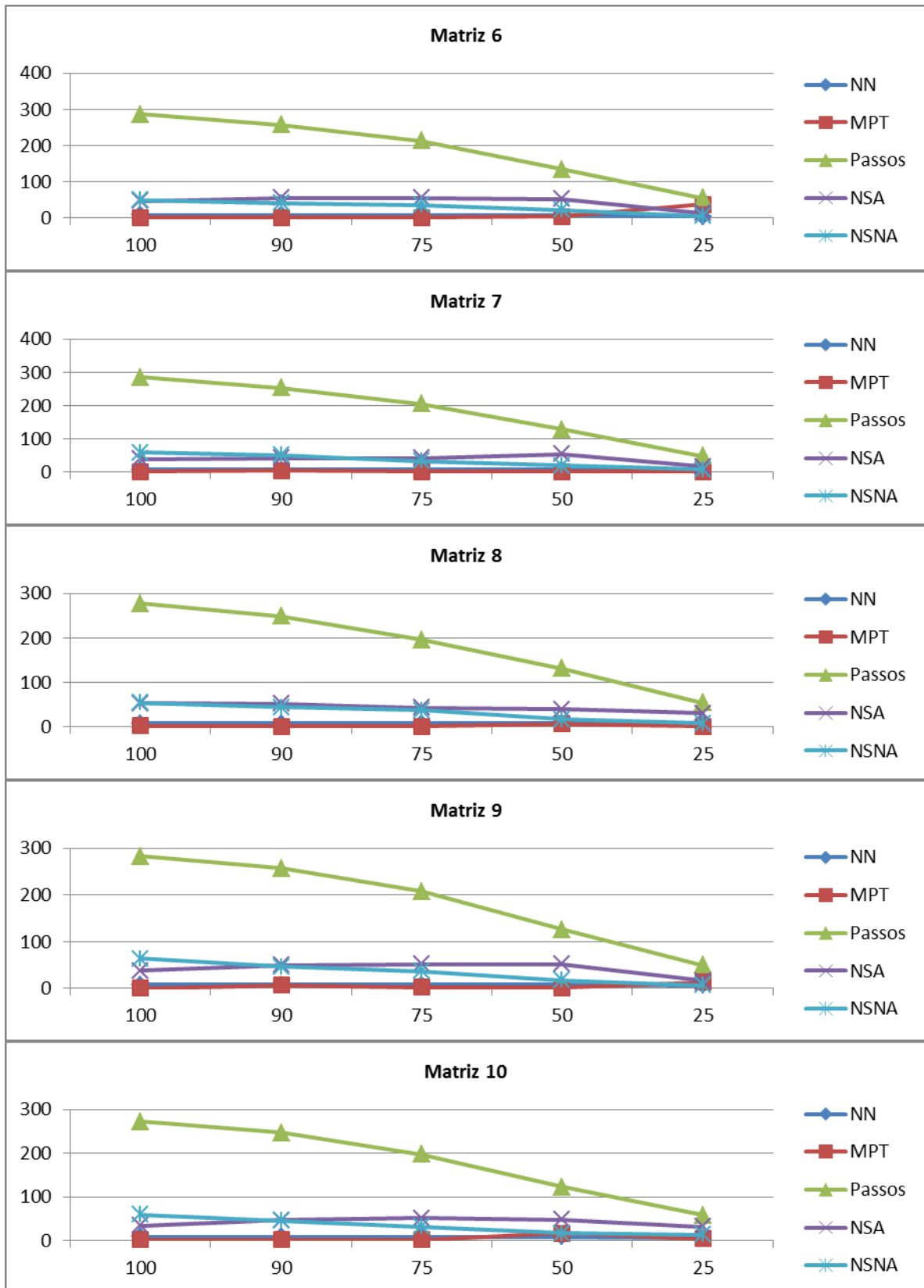
ROWE, T 1988. Definition, diagnosis and origin of Mammalia. **Journal of Vertebrate Paleontology** 8:241-264.

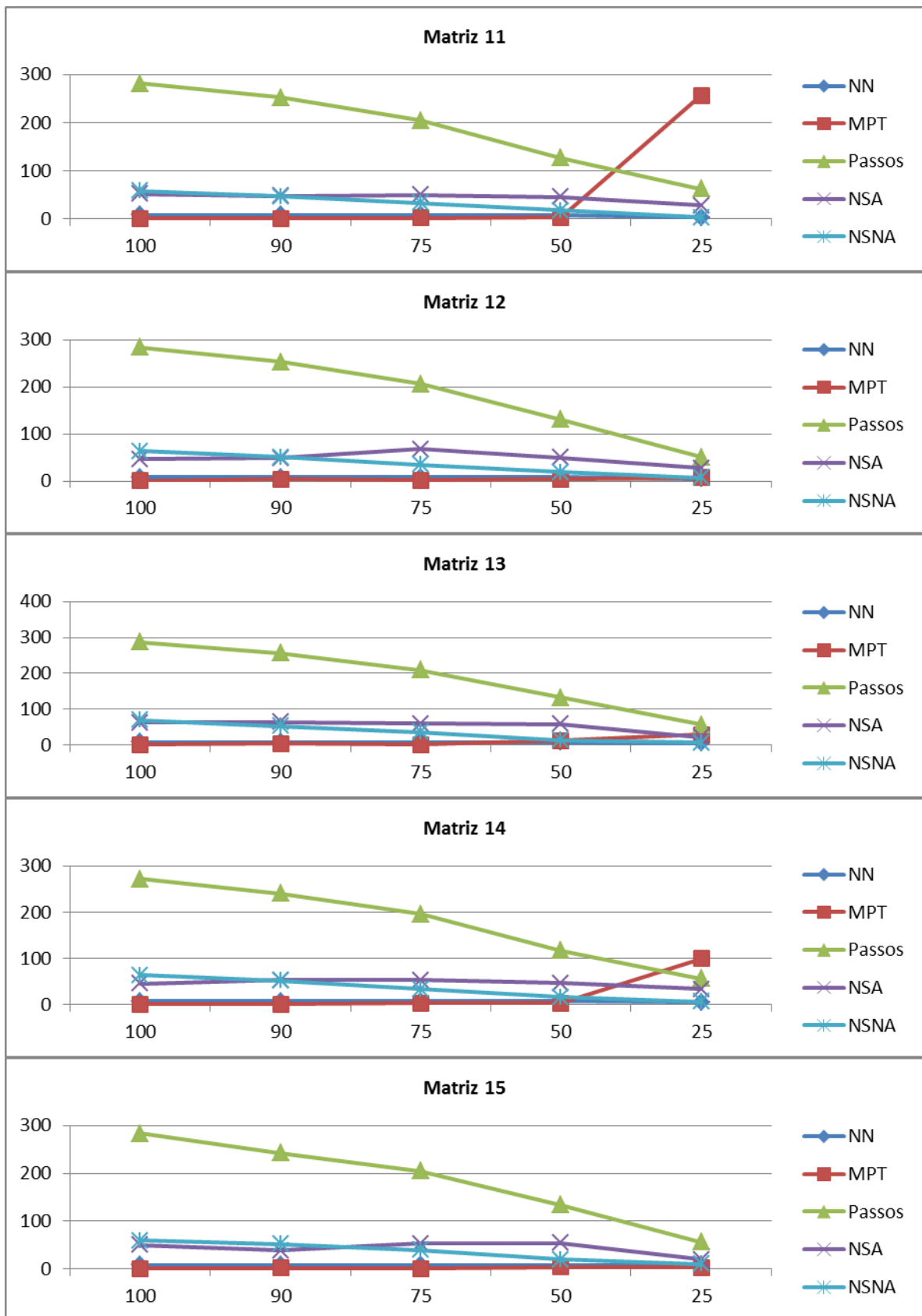
SWOFFORD, D. L. 2001. PAUP*: Phylogenetic Analysis Using Parsimony*, Version 4.0b8. Sinauer, Sunderland, Massachusetts.

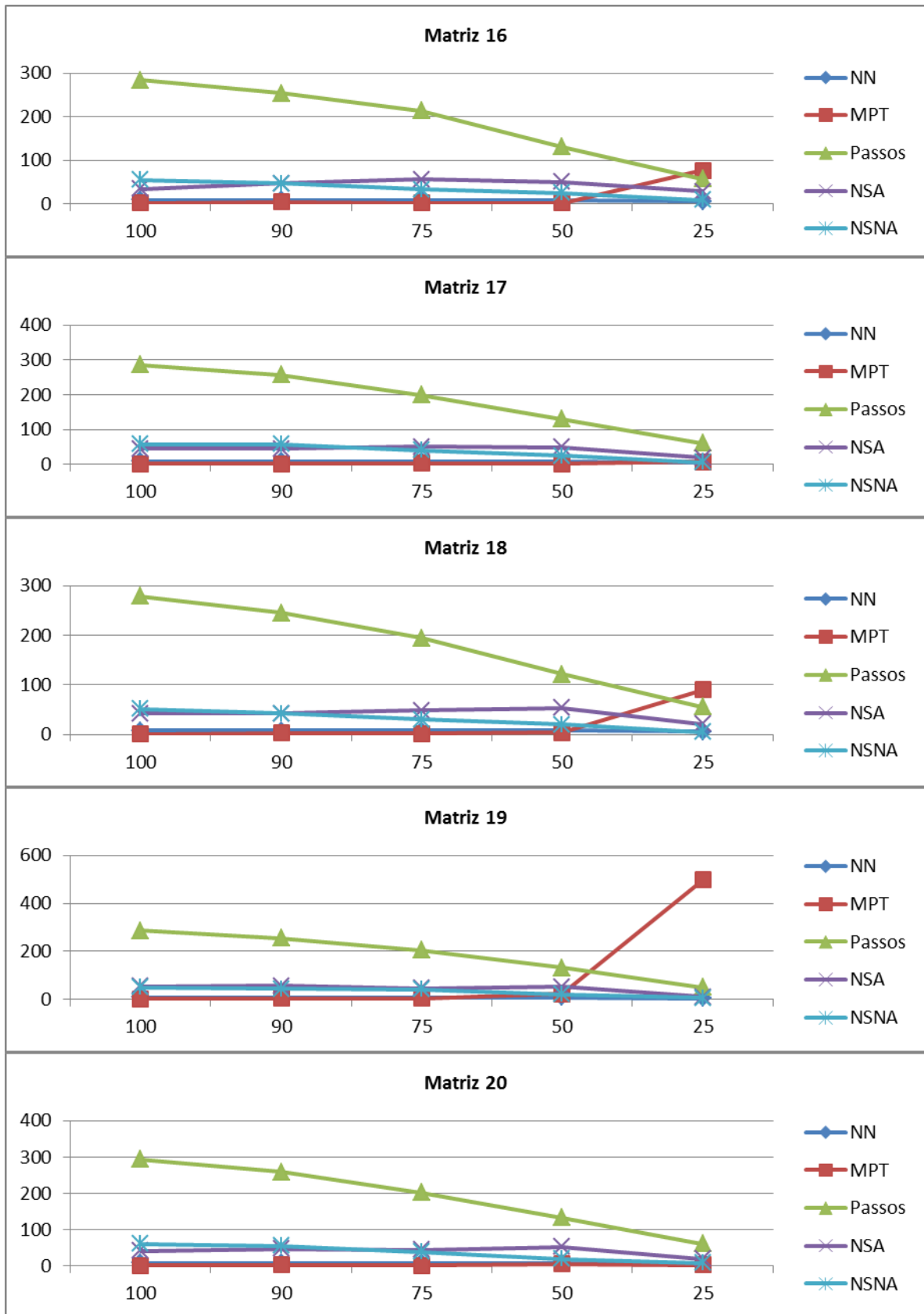
ANEXOS

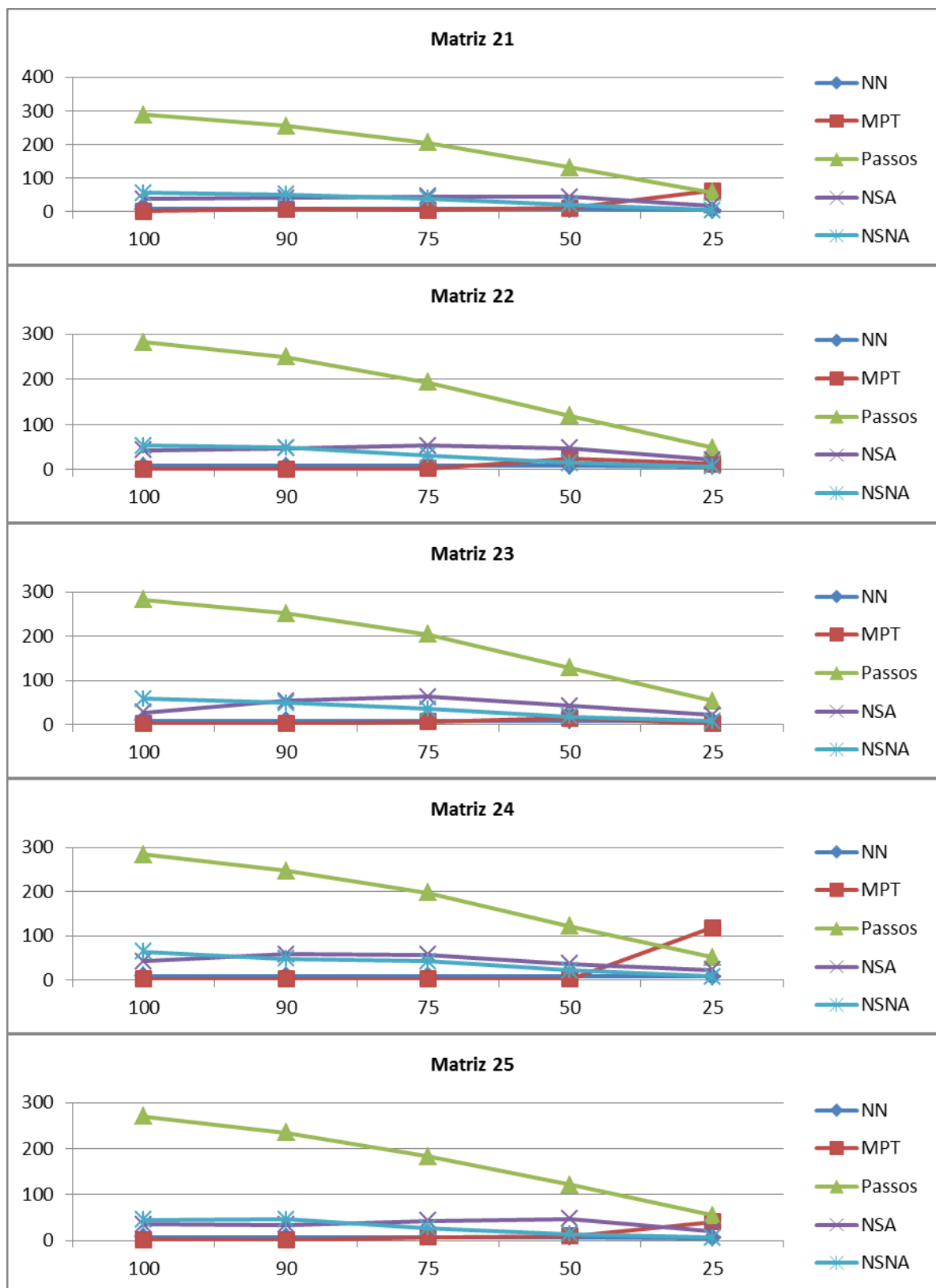
Gráficos das matrizes de 10 táxons/100caracteres.



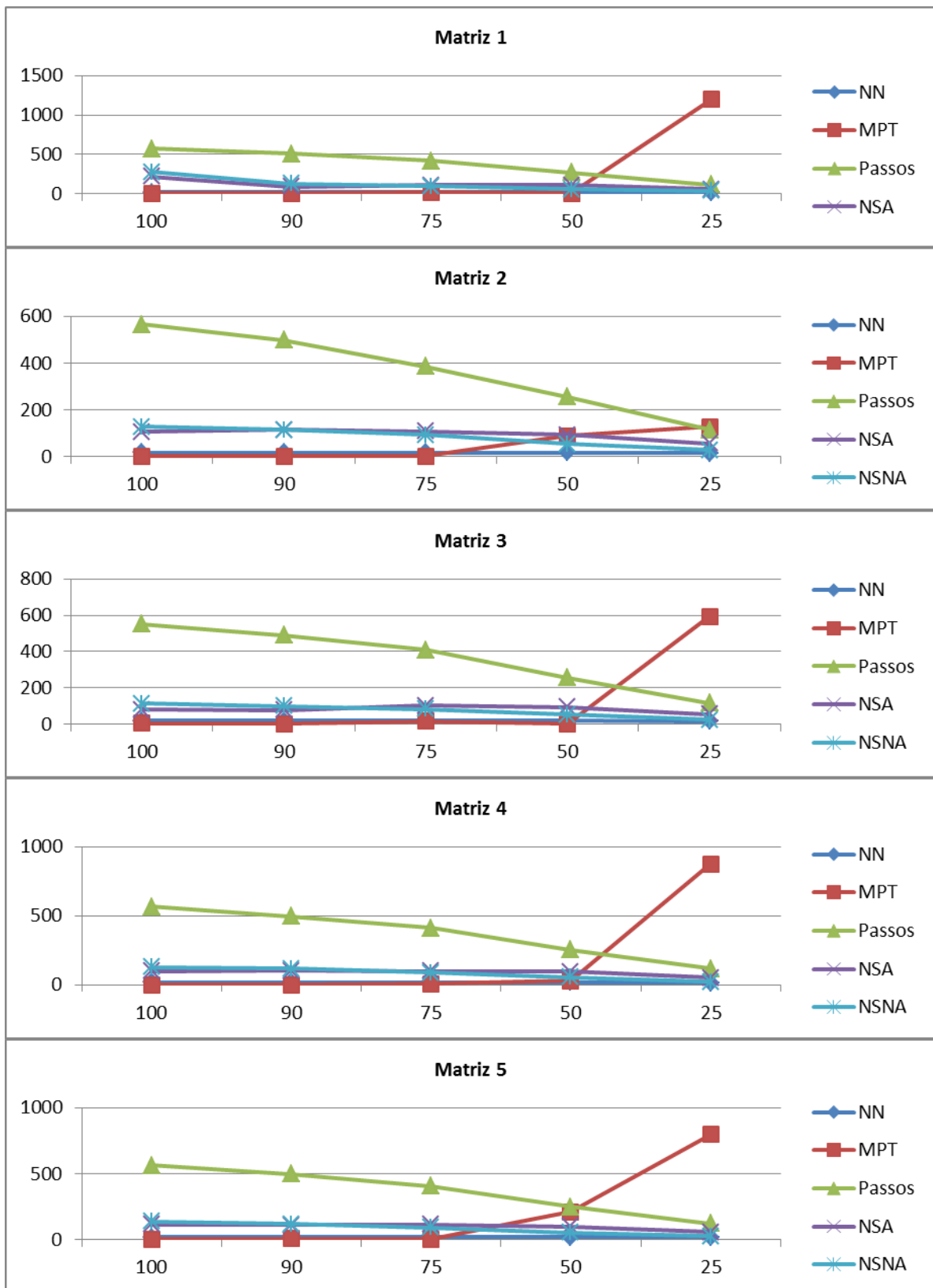


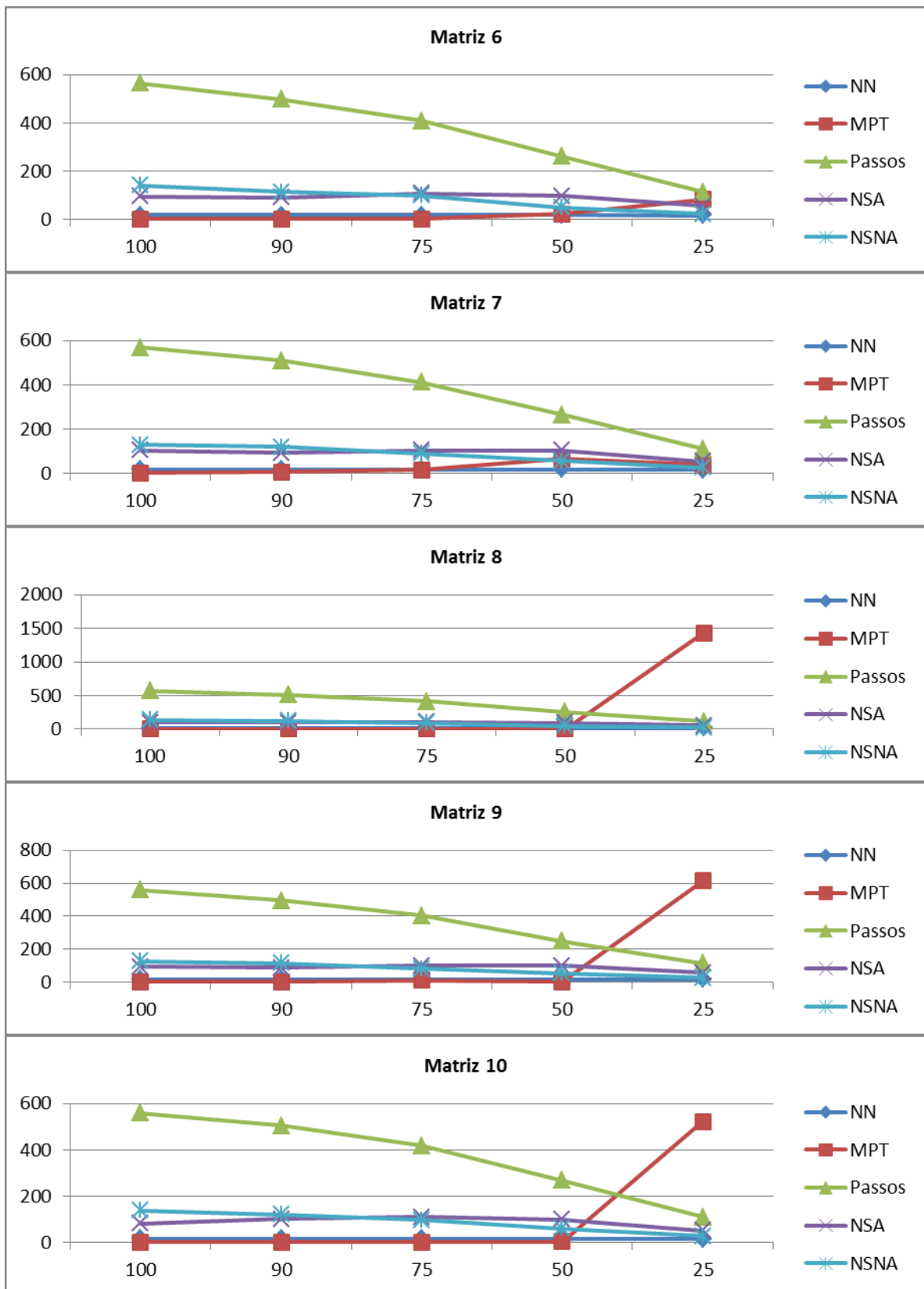


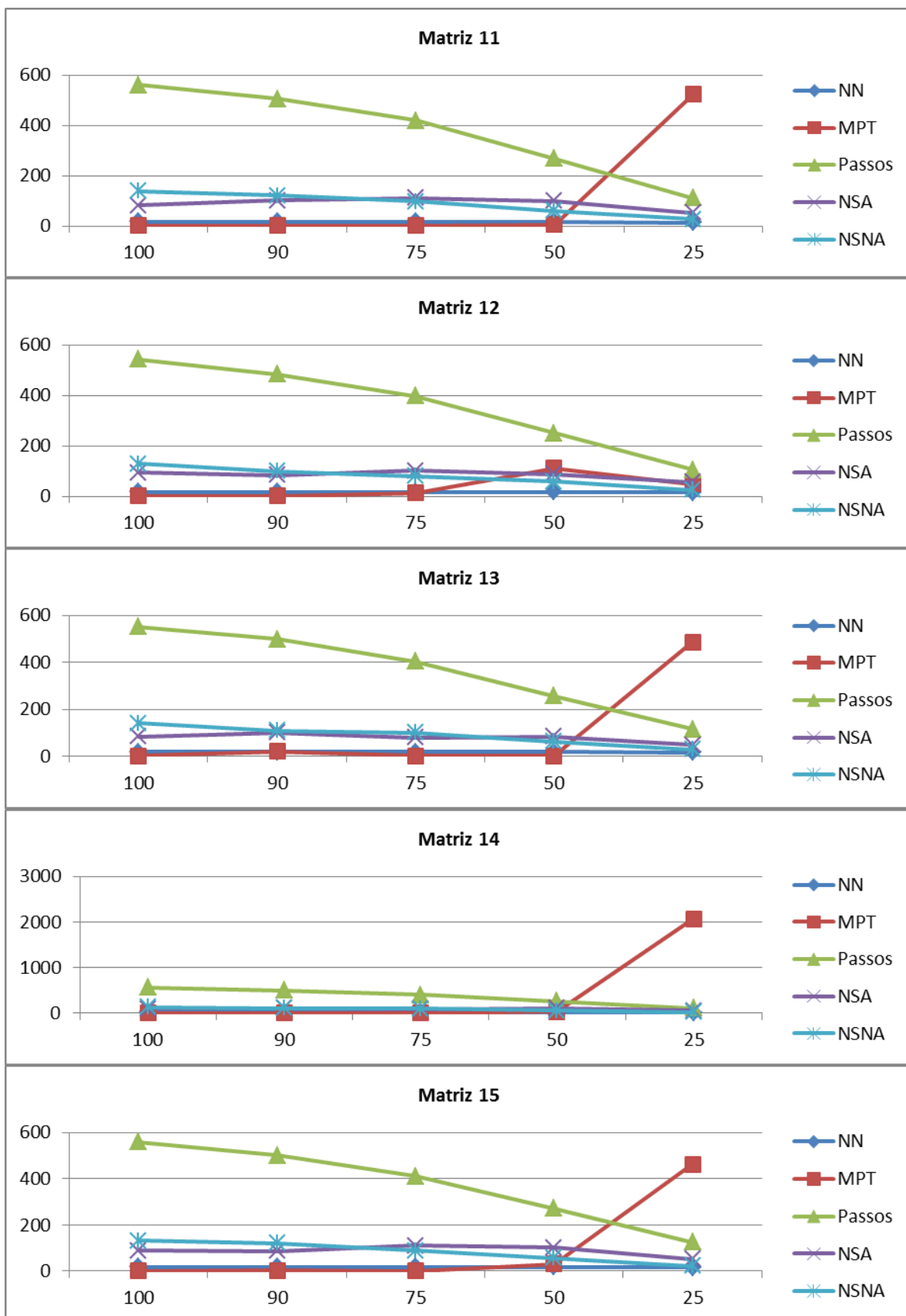


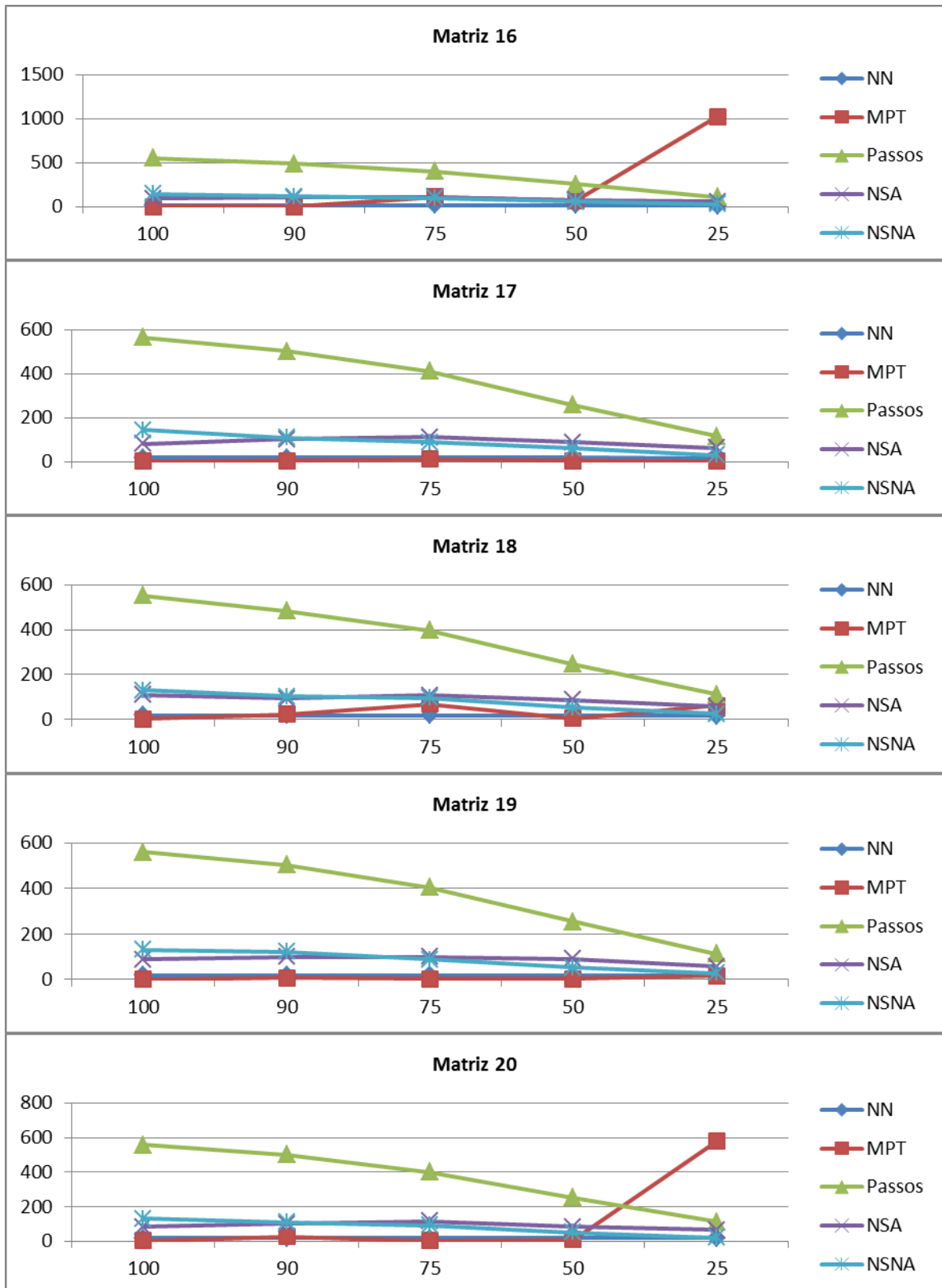


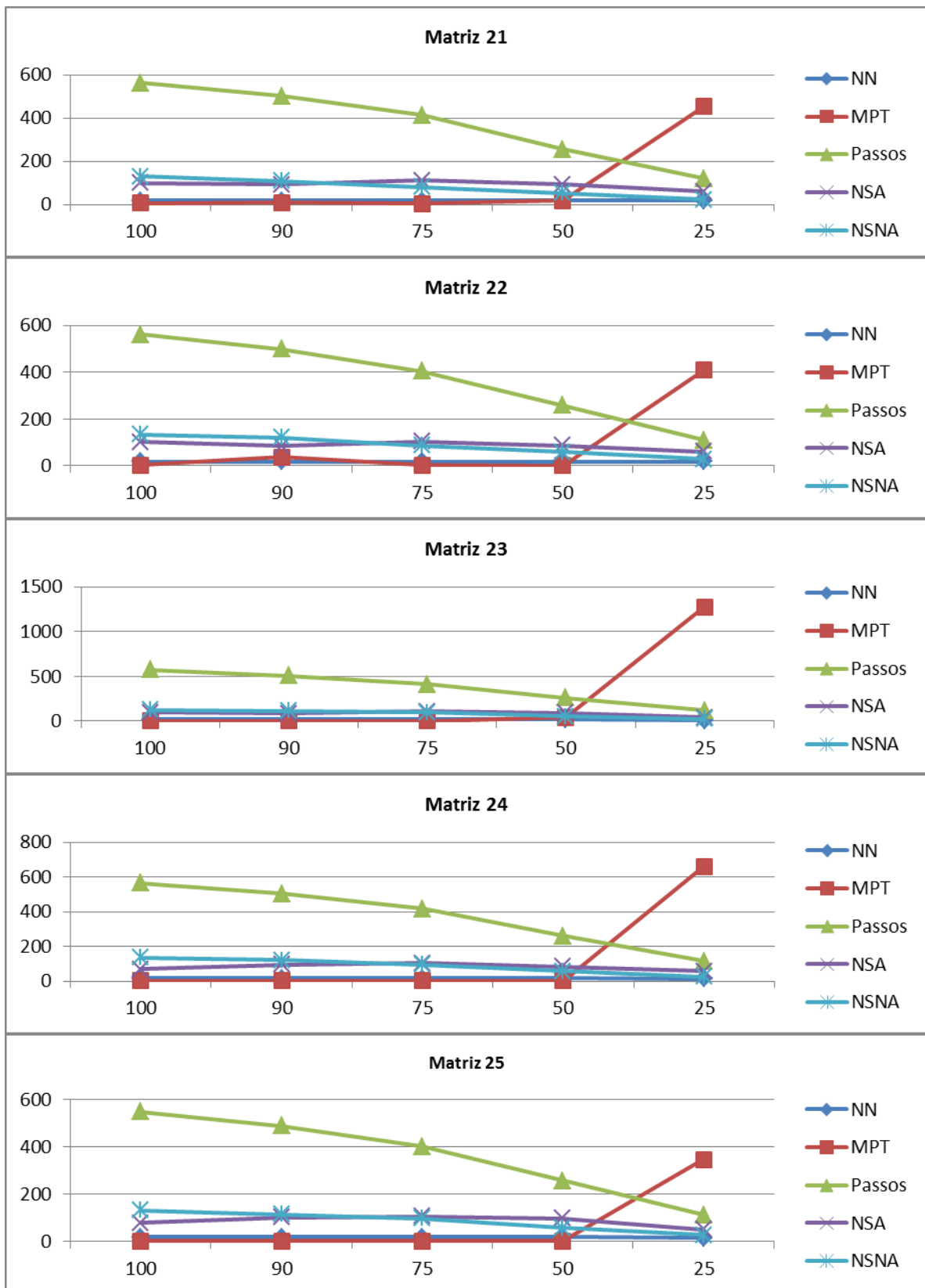
Gráficos das matrizes de 20 táxons/100 caracteres.











Planilha 1 (10 táxons e 100 caracteres)

Matriz	NN 100%	MPT	Passos	NSA	NSNA	NN 90%	MPT	Passos	NSA	NSNA	NN 75%	MPT	Passos	NSA	NSNA	NN 50%	MPT	Passos	NSA	NSNA	NN 25%	MPT	Passos	NSA	NSNA
1	8	2	281	37	49	8	1	246	51	47	8	5	204	47	37	8	6	119	43	16	6	63	52	24	8
2	8	1	281	38	59	8	2	230	58	50	8	5	209	56	36	8	1	126	44	20	7	5	60	16	9
3	8	1	286	49	68	8	2	253	40	47	8	3	204	51	34	8	3	123	35	16	6	94	61	27	6
4	8	2	290	44	57	8	3	263	53	46	8	9	215	60	39	8	1	134	53	21	8	71	53	25	7
5	8	1	285	54	53	8	1	257	37	47	8	2	207	61	38	8	9	131	51	18	7	2	54	29	8
6	8	1	286	47	50	8	2	258	55	42	8	1	213	55	36	8	4	134	52	22	4	37	56	14	6
7	8	1	286	39	59	8	4	254	43	51	8	1	205	43	34	8	2	128	55	20	5	1	49	17	8
8	8	2	278	53	54	8	1	249	52	44	8	1	196	43	38	8	7	131	40	17	8	1	54	31	8
9	8	1	283	38	64	8	7	257	50	47	8	3	208	52	37	8	2	126	52	18	7	13	50	17	7
10	8	1	272	34	59	8	1	247	46	45	8	1	198	51	31	8	16	123	48	16	6	4	58	31	13
11	8	1	281	52	58	8	1	252	48	47	8	2	205	50	33	8	3	127	46	18	3	256	62	28	3
12	8	1	284	47	64	8	4	253	49	51	8	2	206	68	35	8	4	131	50	19	5	7	51	27	7
13	8	1	287	62	69	8	3	256	64	53	8	1	209	59	36	8	12	132	58	14	5	29	57	22	6
14	8	1	272	45	64	8	1	241	53	52	8	3	196	53	33	8	3	117	47	16	5	100	56	34	6
15	8	1	283	50	59	8	2	242	38	52	8	1	205	53	39	8	4	134	54	19	7	3	56	19	10
16	8	1	284	34	54	8	4	254	47	46	8	2	214	56	33	8	2	131	50	24	5	77	57	28	8
17	8	1	286	44	58	8	1	257	45	58	8	3	200	50	40	8	2	130	49	24	6	7	60	18	6
18	8	1	279	43	52	8	3	245	43	43	8	2	195	49	31	8	4	122	54	20	6	90	56	21	5
19	8	1	287	53	50	8	3	255	56	44	8	3	205	46	41	8	22	132	51	20	3	498	49	9	6
20	8	1	294	42	61	8	3	259	47	55	8	1	202	45	39	8	6	133	53	20	6	3	61	20	8
21	8	1	288	38	56	8	7	256	41	52	8	5	205	46	40	8	11	132	44	21	4	62	57	18	5
22	8	1	282	43	53	8	2	250	48	49	8	3	194	53	32	8	25	119	47	15	7	13	49	22	7
23	8	1	283	27	58	8	2	252	53	49	8	6	205	63	35	8	14	128	42	18	7	1	53	21	7
24	8	1	284	42	64	8	2	247	58	48	8	2	197	57	42	8	1	121	36	21	7	118	51	22	7
25	8	2	270	35	45	8	1	235	34	47	8	8	183	43	28	8	10	122	48	15	6	41	55	21	7

Planilha 2 (20 táxons e 100 caracteres)

Matriz	NN 100%	MPT	Passos	NSA	NSNA	NN 90%	MPT	Passos	NSA	NSNA	NN 75%	MPT	Passos	NSA	NSNA	NN 50%	MPT	Passos	NSA	NSNA	NN 25%	MPT	Passos	NSA	NSNA
1	18	1	569	214	272	18	4	505	85	123	18	17	414	104	90	18	1	265	103	57	15	1199	115	54	32
2	18	2	566	107	128	18	2	498	115	115	18	1	386	108	93	18	88	256	93	54	15	127	115	57	29
3	18	3	552	79	112	18	1	489	78	100	18	16	408	102	83	18	1	256	95	52	16	594	115	54	25
4	18	2	566	99	129	18	1	495	103	119	18	8	412	101	88	18	29	256	97	55	14	874	117	53	24
5	18	2	561	113	136	18	4	496	107	115	18	1	404	111	88	18	207	249	97	52	15	797	119	54	24
6	18	2	564	96	141	18	2	498	91	115	18	2	409	109	99	18	23	262	97	49	16	83	114	56	23
7	18	2	569	104	129	18	7	509	94	121	18	17	412	104	90	18	66	264	104	56	15	38	111	54	28
8	18	1	568	98	134	18	2	508	102	122	18	5	412	110	91	18	2	260	95	51	17	1430	115	53	23
9	18	1	560	94	128	18	2	497	88	116	18	10	404	98	82	18	2	249	98	55	16	614	115	58	26
10	18	2	560	82	138	18	2	505	102	121	18	2	419	111	98	18	5	269	100	60	14	523	111	52	27
11	18	1	564	99	130	18	1	499	91	112	18	4	401	107	93	18	98	259	93	53	13	414	115	52	23
12	18	1	543	94	129	18	2	484	84	99	18	13	399	101	81	17	112	251	89	58	15	49	106	56	25
13	18	1	552	83	140	18	19	499	99	109	18	1	404	79	100	18	1	256	82	62	15	487	114	47	28
14	18	1	559	95	128	18	1	500	104	113	18	1	400	90	96	18	34	255	104	53	14	2067	110	59	25
15	18	1	560	89	131	18	3	502	87	121	18	1	412	110	90	18	32	271	101	57	15	462	125	52	23
16	18	1	553	91	143	18	1	487	101	116	18	108	401	108	95	17	66	254	72	58	13	1023	111	55	27
17	18	1	566	82	145	18	4	502	102	110	18	12	411	112	87	18	3	257	89	61	16	1	116	63	29
18	18	1	552	108	130	18	21	483	92	101	18	67	397	106	94	17	3	247	85	52	16	60	110	56	24
19	18	1	559	89	130	18	7	503	100	122	18	2	406	100	89	18	2	255	90	53	16	17	114	57	27
20	18	2	557	82	129	18	27	500	100	110	18	4	399	115	91	18	9	252	84	49	16	578	113	66	21
21	18	6	564	99	131	18	9	502	92	110	18	3	415	112	82	18	19	256	93	52	17	454	120	60	23
22	18	3	561	102	134	18	37	500	84	121	18	3	405	105	87	18	1	258	87	59	18	409	110	62	28
23	18	1	573	98	125	18	3	508	84	116	18	1	413	109	97	18	43	261	93	55	10	1276	119	41	26
24	18	1	565	69	136	18	1	504	96	120	18	2	418	103	95	18	2	259	82	56	14	661	115	58	24
25	18	1	548	78	132	18	1	489	103	114	18	2	403	106	98	18	2	257	97	57	14	346	113	52	27

